B10b Mathematical Models of Financial Derivatives
Hilary Term 2013

Michael Monoyios
Mathematical Institute
University of Oxford

January 17, 2013

Useful books

The books by Shreve [12, 13] are excellent on (respectively) probabilisitic aspects of the binomial model and on stochastic calculus for finance. Etheridge [4] is a good stochastic calculus primer for finance, Björk [1] covers many finance topics outside the scope of the course, Wilmott et al [15] is good on the PDE aspects of the subject, and background on financial derivatives is given in Hull [7]. Jacod and Protter [8] and Grimmett and Stirzaker [5] are useful for background probability material.

The lecture notes and background material

These notes contain the core material. Some material is marked with an asterix and is not examinable. Some probability theory underlying conditional expectation and martingales is contained in the supplementary notes *Background Probability*, available on the course website. These are for those who wish to brush up on some probabilistic material, and are (I hope) helpful in developing intuition for notions like filtrations and adaptedness of stochastic proicesses, using the binomial stock price model as an example.

Below are some comments on the level of mathematical knowledge you will need to acquire. Its application to finance, as exemplified by the binomila modle and the Black-Scholes model, is fundamental to the course.

The Background Probability material is not examinable. You are expected to become familiar with the use of some probabilistic termninology (σ-algebras, filtrations, random variables that are measurable with respect to a σ-algebra). You are expected to have some familiarity with the properties of conditional expectation (you will not be examined on proofs of these) and martingales, and to be able to use them.

You are expected to know the defining properties of a stochastic process $B = (B_t)_{t\geq 0}$ known as *Brownian motion* (BM), and to understand how these lead to the fact that its *quadratic variation* (QV) process $[B]$ is equal to the time elapsed: $[B]_t = t$. You are expected to know (but not prove) Lévy's criterion: any continuos martingale M satisfying $[M]_t = t$ is a BM.

You should be able to use the properties of Brownian motion (such as its independent Gaussian increments property and its quadratic variation property) and you should have some appreciation of how these lead to the properties of the Itô integral (such as the martingale property and the Itô isometry) for elementary integrands (though you are not required to know the full theory of the construction of the Itô integral for general integrands).

You are expected to have an appreciation of how the quadratic variation property leads to the Itô formula and to properties of the Itô integral (that is, *stochastic calculus*). You are expected to be able to use the *Itô formula* (both the one-dimensional and multi-dimensional versions) fluently. You are expected to understand (and prove using the Itô formula) the connection between PDEs and stochastic calculus, in the form of the Feynman-Kac theorem.

Contents

Part I
Introduction to derivative securities

1 Financial derivatives

Definition 1.1. A *European derivative security* (or *European contingent claim*) is a financial contract which pays its holder a random amount (the *payoff* of the claim) at some future time T (the *maturity time* of the derivative).

An *American* derivative delivers the payoff at a random time $\tau \leq T$ chosen by the holder of the contract. The payoff is (typically) contingent on the value of some other underlying security (or securities), or on the level of some non-traded reference index.

Example 1.2 (Forward contract). The holder of a forward contract agrees to buy an asset at some future time T for a fixed price K (the *delivery price*) that is decided at initiation of the contract. Hence, the forward contract has a value (to the holder) at maturity of $S_T - K$, where S_T is the underlying asset value at maturity.

The forward contract thus allows the holder to fix the purchase price of the underlying asset in advance, and so can be used to mitigate the risk inherent in the price uncertainty (that is, to *hedge* the price risk). It can also be used to speculate against future price moves.

Example 1.3 (European call option). A European call option has payoff $(S_T - K)^+$ at maturity. It confers to the owner the right (but not the obligation) to buy the underlying asset at maturity time T for a fixed price K (the *strike price*, or *exercise price*).

The origins of derivatives lie in medieval agreements between farmers and merchants to trade the farmer's harvest at some future date, at a price set in advance. This allowed farmers to fix the selling price of their crop, and reduced the risk of having to sell at a lower price than their cost of production, which might happen in a bumper harvest year. This is one motivation for the existence of derivatives: they give random payoffs which can be used to eliminate uncertainty from future asset price trades. The act of removing uncertainty in finance is called *hedging*.

Consider a farmer whose unit cost of crop production is C. His profit on selling the crop would be $S_T - C$, where S_T is the market crop price at harvest time. If the farmer were to sell (that is, take a *short* position in) a forward contract with delivery price $K > C$, at some time $t < T$, then his overall payoff at T would be $S_T - C - (S_T - K) = K - C > 0$. The risk of the crop price being less than the cost of production has been removed.

Since derivatives have random payoffs, they can also be used to take risk by speculating on the future values of asset prices, and they are often a cheaper device for doing so than investing in the underlying asset. For instance, a European call option on a stock, with payoff $(S_T - K)^+$, where S_T is the (random) stock price at time T and $K \geq 0$ is a constant called the *strike price* of the option, allows the holder of the call to take a positve payoff if the stock price is above K, and the cost of acquiring a call option is usually only a fraction of the cost of buying the stock itself.

This course will be about how to assign a value to a derivative at any time $t \leq T$. This will involve modelling the randomness in the underlying asset price process $S = (S_t)_{0 \leq t \leq T}$.

To do this, we will need the notion of a *stochastic process* on a *filtered probability space.* We shall see that the key to valuing derivatives is to attempt to use the underlying asset to remove the risk from selling (or buying) the derivatieve. That is, derivative valuation is via a hedging argument.

1.1 Underlying assets

Typical assets which are traded in financial markets, and which can be the underlying assets for a derivative contract, include:

- shares (stocks)
- commodities (metals, oil, other physical products)
- currencies
- bonds (assets used as borrowing tools by govenments and companies) which pay fixed amounts at regular intervals to the bond holder.

An agent who holds an asset will be said to hold a *long position* in the asset, or to be *long* in the asset.

An agent who has sold an asset will be said to hold a *short position* in the asset, or to be *short* in the asset.

For the most part in this course, we will focus on derivative securities which have a *stock* as underlying asset. The stock price will be a stochastic process denoted by $S = (S_t)_{0\leq t\leq T}$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$. This means that for each $t \in [0, T]$, S_t, the value of the stock at time t, is a random variable that is measurable with respect to the σ-algebra $\mathcal{F}_t$. That is, S is a process that is *adapted* to the filtration $\mathbb{F}$. We shall see later that this means the following: each $\mathcal{F}_t$ is a collection of subsets of a set Ω (the sample space), closed under complements and under countable unions, and with $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$ (such an increasing sequence of σ-algebras is called a *filtration*, and will represent increasing information as time evolves). Each S_t is a function from Ω to $\mathbb{R}^+$ with the property that sets of the form

$$\{\omega \in \Omega | S_t(\omega) \in A \subseteq \mathbb{R}^+\}$$

lie in $\mathcal{F}_t$. This is what we mean by saying that S_t is an $\mathcal{F}_t$-measurable random variable. The adaptedness property (S_t is $\mathcal{F}_t$-measurable for each $t \in [0, T]$) is tantamount to the idea that the information available at time t is sufficient to know the value of S_t (that is, if you observe the stock market up to time t, you will know the current value of the stock price).

1.2 Interest rates and time value of money

Let us measure time in some convenient units, say years. If an interest rate r is quoted per annum and with compounding frequency at time intervals δt, this means that an amount A invested for a time period δt will grow to $A(1 + r\,\delta t)$. If this is re-invested for another period δt, the balance becomes $A(1 + r\,\delta t)^2$, and so on. So after n periods, with $t := n\,\delta t$, we have

$A(1 + r\,\delta t)^n = A(1 + rt/n)^n$. A *continuously compounded* interest rate corresponds to the limit $n \to \infty$, or $\delta t \to 0$. In this case, after time t an amount A will grow to

$$\lim_{n\to\infty} A\left(1 + \frac{rt}{n}\right)^n = Ae^{rt}.$$

So an amount A invested at time zero for a time t will grow to an amount Ae^{rt}, where r is the continuously compounded risk-free interest rate. We call the amount Ae^{rt} the *future value* of A invested at time zero, and the factor e^{rt} is called an *accumulation factor*.

By the same token, receiving an amount A at time t is equivalent to receiving Ae^{-rt} at time zero. We call Ae^{-rt} the *present value* of A received at time t (we say that A is *discounted* to the present) and the factor e^{-rt} is called a *discount factor*.

It is usually convenient (but nothing more) to assume that interest is continuously compounded. We do not need to assert that, in reality, interest is continuously compounded, in order to use a continuously compounded interest rate in all our analysis. If the interest is actually compounded m times a year at an interest rate of R per annum, then we can still use a continously compounded interest rate r simply by making the identification

$$A\left(1 + \frac{R}{m}\right)^{mt} = A\exp(rt), \tag{1.1}$$

so that there is a one-to-one correspondence between the interest rate R (compounded m times per annum) and the continuously compounded interest rate r. In this course, we will nearly always use continuously compounded rates when considering continuous time models.

A differential version of the above arguments is as follows. In continuous time, we model the time evolution of cash in a bank account in terms of a riskless asset which we shall call a *money market account*, which is the value at time $t > 0$ of \$1 invested at time zero and continuously earning interest which is reinvested. We shall denote the value of this asset at time t by B_t, which satisfies

$$\mathrm{d}B_t = rB_t\,\mathrm{d}t, \quad B_0 = 1, \tag{1.2}$$

where r is the (assumed constant) interest rate. Then the value of the bank account at time t is given by $B_t = e^{rt}$, which we see is the accumulation factor we encountered above.

A more complex model could assume that interest rates are time-varying (possibly stochastic). In this case the money market account would satisfy

$$\mathrm{d}B_t = r_tB_t\,\mathrm{d}t, \quad B_0 = 1, \tag{1.3}$$

where r_t is the instantaneous (or short term) interest rate. We have allowed for this to be time-varying, and r_t represents the interest rate in the time interval $[t, t + dt)$. From (1.3) we see that

$$B_t = \exp\left(\int_0^t r_u\,\mathrm{d}u\right), \tag{1.4}$$

and this is the accumulation factor in this case. This is the factor by which \$1 invested at time zero grows to at time t, when the interest generated is continually reinvested.

1.3 Forwards and futures

Definition 1.4 (Forward contract)**.** A forward contract obliges its holder to buy an underlying asset (a stock, say) at some future time T (the *maturity time*) for a price K (the *delivery price* that is fixed at contract initiation. Hence, at time T, when the stock price is S_T, the contract is worth $S_T - K$ (the *payoff* of the forward) to the holder. This payoff is shown in Figure 1.

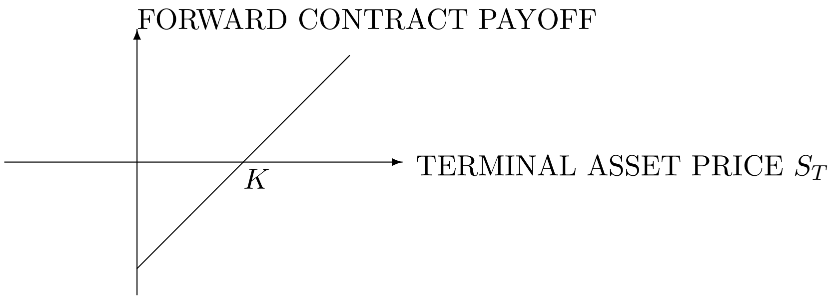


Figure 1: Forward contract payoff as function of final underlying asset price

A *futures contract* is a rather specialised forward contract, traded on an organised exchange, and such that, if a contract is traded at some time $t \leq T$, the delivery price is set to a special value $F_{t,T}$, called the *futures price of the asset* or the *forward price of the asset*, chosen so that the value of the futures contract at initiation (that is, at time t), is zero.

Futures markets have other specialised features that we will encounter later. Principally, a participant in a futres market is required to set up a so-called *margin account* as collateral, so that one's daily profits and losses are reflected by adjustments in the margin account. One also has to maintain the balance in the margin account at some minimum value (the maintenance margin), and recives a so-called *margin call* (a demand to top-up the margin account) if the balance in the margin account falls below the maintenance margin. This mechanism is designed to remove the risk of defualt from the market, and hence futures markets are very liquid. One can consult Hull [7] for detailed descriptions of the workings of futures excahnges.

1.3.1 Valuation of forward contracts

In what follows we value forward contracts on a non-dividend paying stock, that is, an asset with price process $S = (S_t)_{0 \leq t \leq T}$ that pays no income to its holder.

Lemma 1.5. *The value at time $t \leq T$ of a forward contract with delivery price K and maturity T, on an asset with price process $S = (S_t)_{0 \leq t \leq T}$, is $f_{t,T} \equiv f(t, S_t) \equiv f(t, S_t; T) \equiv f(t, S_t; K, T)$, given by*

$$f_{t,T} = S_t - K \exp(-r(T-t)), \quad 0 \leq t \leq T. \tag{1.5}$$

Proof. This is a simple hedging argument which provides our first example of a riskless hedging strategy. Start with zero wealth at time $t \leq T$, and sell the contract at time $t \leq T$ for some price $f_{t,T}$. Hedge this sale by purchasing the asset for price S_t. This requires borrowing of $S_t - f_{t,T}$.

At time T, sell the asset for price K under the terms of the forward contract, and require that this is enough to pay back the loan. Hence we must have

$$K = (S_t - f_{t,T}) \exp(r(T-t)),$$

and the result follows.

□

Corollary 1.6. *The forward price of the asset at time $t \leq T$ is given by*

$$F_{t,T} = S_t \exp(r(T-t)), \quad 0 \leq t \leq T.$$

Proof. Set $f_{t,T} = 0$ in Lemma 1.5 and then by definition we must have $K = F_{t,T}$.

□

1.4 Arbitrage

The simple argument above for valuing a forward contract is an example of valuation by the principle of *no arbitrage*. If the relationship in Lemma 1.5 is violated, then an elementary example of a riskless profit opportunity, called an *arbitrage*, ensues. Here is a definition of arbitrage.

Definition 1.7 (Arbitrage)**.** Let $X = (X_t)_{0 \leq t \leq T}$ denote the wealth process of a trading strategy. An *arbitrage* over $[0, T]$ is a strategy satisfying $X_0 = 0$, $\mathbb{P}[X_T \geq 0] = 1$ and $\mathbb{P}[X_T > 0] > 0$.

So an arbitrage is guaranteed not to lose money and has a positive probability of making a profit. If the valuation formula (1.5) for the forward contract is violated, an immediate arbitrage opportunity occurs, as we now illustrate.

Suppose $f_{t,T} > S_t - K \exp(-r(T-t))$. Then one can short the forward contract and buy the stock, by borrowing $S_t - f_{t,T}$ at time t. At maturity, one sells the stock for K under the terms of the forward contract and uses the proceeds to pay back the loan, yielding a profit of

$$K - (S_t - f_{t,T}) \exp(r(T-t)) > 0.$$

This is an arbitrage. A symmetrical argument applies if $f_{t,T} < S_t - K \exp(-r(T-t))$ (and you should supply this).

The principle of riskless hedging and no arbitrage will also apply, rather less trivially, to the valuation of options later in the course.

An equivalent way of looking at no arbitrage is sometimes called the *law of one price*. Two portfolios which give the same payoff at T should have the same value at time $t \leq T$. Let us show how this applies to the valuation of a forward contract. Consider the following two portfolios at time $t \leq T$:

- a long position in one forward contract,
- a long position in the stock plus a short cash position of $K\exp(-r(T-t))$.

At time T, these are both worth $S_T - K$, so their values at time $t \le T$ must be equal, yielding $f_{t,T} = S_t - K\exp(-r(T-t))$, as before. Notice that the second portfolio *perfectly replicates* (or perfectly hedges) the payoff of the forward contract, meaning that it reproduces the payoff $S_T - K$. Denote the position in the stock that is needed to perfectly hedge a forward contract by H_t^f. Then we have that $H_t^f = 1$ for all $t \in [0,T]$, and note that

$$H_t^f = f_x(t, S_t) = 1, \quad 0 \le t \le T,$$

where $f(t,x) := x - Ke^{-r(T-t)}$. This is a simple example of a "delta hedging rule", in which one differentiates the pricing function of the derivative with respect to the variable representing the underlying asset price, in order to get the hedging strategy. We will see a similar result when valuing options.

1.4.1 Forward on dividend-paying stock

The stock in the preceding analysis was assumed to pay no dividends. Now assume that the stock pays dividends as a continuous income stream with *dividend yield* q. This means that in the interval $[t, t + \mathrm{d}t)$, the income received by someone holding the stock will be $qS_t\,\mathrm{d}t$. Suppose that at time $t \in [0,T]$ an agent holds n_t shares of stock. The income received in the next infinitesimal time interval is $qn_tS_t\,\mathrm{d}t$. If this is immediately re-invested in shares, the holding in shares satisfies

$$\mathrm{d}n_t = qn_t\,\mathrm{d}t.$$

Hence, if the initial holding is n_0 at time zero, we have

$$n_t = n_0\exp(qt), \quad 0 \le t \le T.$$

In particular, we have

$$n_t = n_T\exp(-q(T-t), \quad 0 \le t \le T.$$

This means that in order to hold one share of stock at time T, one may buy $\exp(-q(T-t))$ shares at $t \le T$ and re-invest the dividends in the stock.

If we use this to value a forward contract on the dividend-paying stock we arrive at the following.

Lemma 1.8. *The value at time $t \le T$ of a forward contract with delivery price K and maturity T, on a stock with price process $S = (S_t)_{0\le t\le T}$ paying dividends at a dividend yield q, is given by*

$$f_{t,T} = S_t\exp(-q(T-t)) - K\exp(-r(T-t)), \quad 0 \le t \le T. \tag{1.6}$$

Proof. This is again a hedging argument. Start with zero wealth at time $t \le T$, and sell the contract at time $t \le T$ for some price $f_{t,T}$. Hedge this sale by purchasing $\exp(-q(T-t))$

shares at price S_t. This requires borrowing of $\exp(-q(T-t))S_t - f_{t,T}$. Re-invest all dividends immediately in the stock, so as to hold one share at time T.

At time T, sell the asset for price K under the terms of the forward contract, and ensure that this is enough to pay back the loan. Hence we must have

$$K = (S_t \exp(-q(T-t)) - f_{t,T}) \exp(r(T-t)),$$

and the result follows. □

Corollary 1.9. *The forward price of the dividend-paying asset at time $t \leq T$ is given by*

$$F_{t,T} = S_t \exp((r-q)(T-t)), \quad 0 \leq t \leq T.$$

Proof. Set $f_{t,T} = 0$ in Lemma 1.8 and then by definition we must have $K = F_{t,T}$. □

Remark 1.10 (Forwards and futures on currencies). A foreign currency is treated as an asset which pays a "dividend yield" equal to the foreign interest rate r_f. Hence, suppose $S = (S_t)_{0\leq t\leq T}$ is the exchange rate (the value in dollars of one unit of foreign currency), then a forward contract on the foreign currency has value at time $t \leq T$ given by

$$f_{t,T} = S_t \exp(-r_f(T-t)) - K \exp(-r(T-t)), \quad 0 \leq t \leq T,$$

where T is the maturity and K is the delivery price.

1.5 Options

An option is a contract that gives the holder the right but not the obligation to buy or sell an asset for some price that is defined in advance.

The two most basic option types are a European call and a European put.

Definition 1.11 (European call option)**.** A *European call option* on a stock is a contract that gives its holder the right (but not the obligation) to purchase the stock at some future time T (the *maturity time*) for a price K (the *strike price* or *exercise price*) that is fixed at contract initiation. If $S = (S_t)_{0\leq t\leq T}$ denotes the underlying asset's price process, the payoff of a call option is $(S_T - K)^+$, as shown in Figure 2.

Definition 1.12 (European put option)**.** A European put option on a stock is a contract that entitles the holder to sell the underlying stock for a fixed price K, the strike price, at a future time T. If $S = (S_t)_{0\leq t\leq T}$ denotes the underlying asset's price process, the payoff of a put option is $(K - S_T)^+$, as shown in Figure 3.

The act of choosing to buy or sell the asset under the terms of the option contract is called *exercising* the option.

Options which can be exercised any time before the maturity date T are called *American* options, whilst *European* options can only be exercised at T. Hence, an American call (respectively, put) option allows the holder to buy (respectively, sell) the underlying stock for price K at any time before maturity.

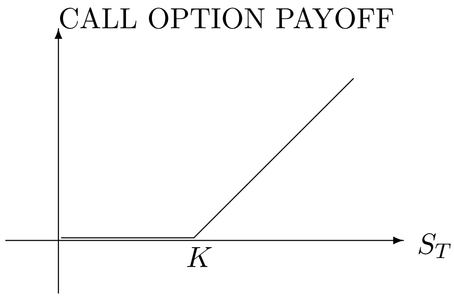


Figure 2: Call option payoff

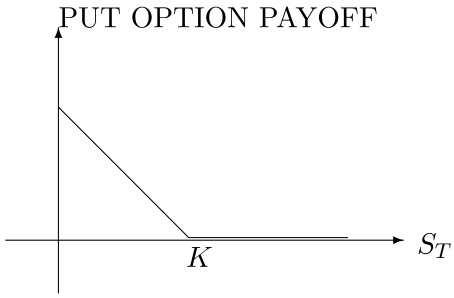


Figure 3: Put option payoff

Lemma 1.13 (Put-call parity). *The European call and put prices $c(t, S_t)$ and $p(t, S_t)$ of options with the same strike K on a non-dividend paying traded stock with price S_t at time $t \in [0, T]$ are related by*

$$c(t, S_t) - p(t, S_t) = S_t - Ke^{-r(T-t)}, \quad 0 \leq t \leq T.$$

Proof. The payoffs of a call and put satisfy

$$c(T, S_T) - P(T, S_T) = (S_T - K)^+ - (K - S_T)^+ = S_T - K,$$

which shows the (obvious) fact that a long position in a call combined with a short position in a put is equivalent to a long position in a forward contract. Hence, their prices at $t \leq T$ must satisfy

$$c(t, S_t) - p(t, S_t) = f_{t,T} = S_t - Ke^{-r(T-t)}, \quad 0 \leq t \leq T,$$

where $f_{t,T}$ is the value of a forward contract at $t \leq T$.

□

Remark 1.14. The same argument applied to a dividend-paying stock yields

$$c(t, S_t) - p(t, S_t) = f_{t.T} = S_t e^{-q(T-t)} - Ke^{-r(T-t)}, \quad 0 \leq t \leq T,$$

where q is the dividend yield.

1.5.1 Europen call and put price bounds

Put-call parity is a model-independent result. From it, we see that $c(t, S_t) \geq f(t, S_t)$. That is, a call option is always at least as valuable as a forward contract (an obvious fact).[1]

A call option gives its holder the right to buy the underlying stock, which means that its value can never be greater than that of the stock, so $c(t, S_t) \leq S_t$.[2] From this we deduce model-independent bounds on a European call option price on a non-dividend-paying stock:

$$S_t - Ke^{-r(T-t)} \leq c(t, S_t) \leq S_t, \quad 0 \leq t \leq T.$$

These bounds are shown in Figure 4.

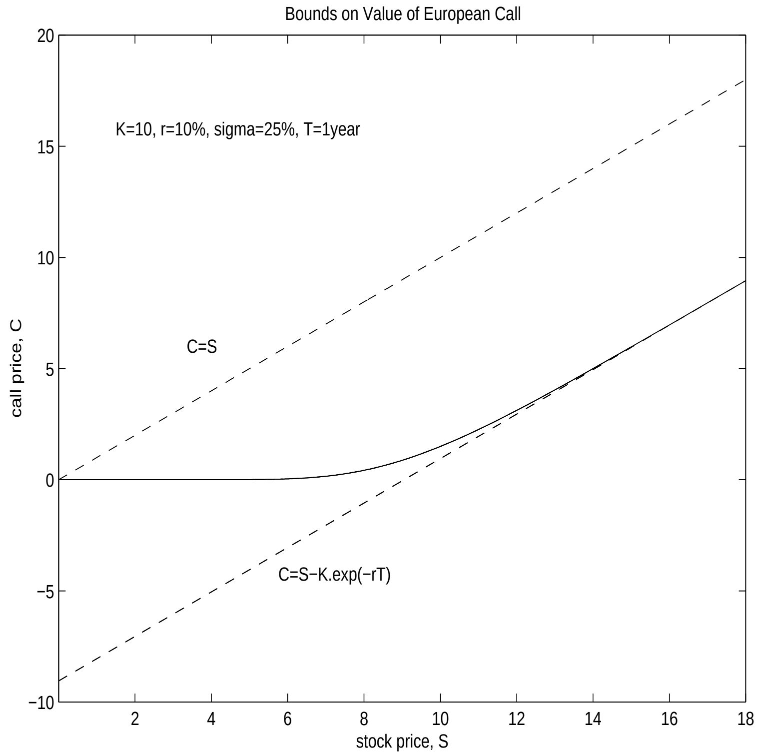


Figure 4: Bounds on European call option value

In Figure 4 we have plotted the upper and lower bounds of a European call value (the dotted graphs) as well as the Black-Scholes value of the call (the solid graph). We will show in these lectures how this function arises. If the above call option pricing bounds are violated, then arbitrage opportunities arise.

[1]Equivalently, comparing the payoff of a call with that of one share plus a short position of $Ke^{-r(T-t)}$ in cash, we obtain $S_T - K \leq (S_T - K)^+$, and therefore $c(t, S_t) \geq S_t - Ke^{-r(T-t)}$.

[2]Equivalently, comparing the payoffs of a share and a call, we have $S_T = S_T^+ > (S_T - K)^+$, and therefore $c(t, S_t) \leq S_t$.

For example, if If $c(t, S_t) < S_t - Ke^{-r(T-t)}$ one should buy the call and short the stock, which gives a cash amount $S_t - c(t, S_t)$ to be invested in a bank account. At time T, we have two possibilities:

1. $S_T \leq K$, in which case the call is not exercised. The arbitrageur buys the stock in the market to close out the short position, using the proceeds from the bank account, which stand at $(S_t - c(t, S_t))e^{r(T-t)}$ prior to buying the stock. This leaves a profit of

$$\begin{aligned}(S_t - c(t, S_t))e^{r(T-t)} - S_T &> (S_t - c(t, S_t))e^{r(T-t)} - K \\ &= e^{r(T-t)}(S_t - Ke^{-r(T-t)} - c(t, S_t)) > 0.\end{aligned}$$

2. $S_T > K$, in which case the call is exercised. The arbitrageur buys the stock for K to close out the short position, using the proceeds from the bank account, which stand at $(S_t - c(t, S_t))e^{r(T-t)}$ prior to buying the stock. This leaves a profit of

$$(S_t - c(t, S_t))e^{r(T-t)} - K = e^{r(T-t)}(S_t - Ke^{-r(T-t)} - c(t, S_t)) > 0.$$

We can derive similar model-independent bounds on a put option price. A put option gives its holder the right to receive an amount K for the stock, so the most it can be worth at maturity is K (if the final stock price is $S_T = 0$). Hence, its current value can never be greater than the present value of K, so that

$$p(t, S_t) \leq Ke^{-r(T-t)}, \quad 0 \leq t \leq T.$$

Similarly, for the value of a put at expiry we have $p(T, S_T) = (K - S_T)^+ \geq K - S_T$. That is, a put option is at least as valuable as a short position in a forward contract. Hence we have the lower bound

$$p(t, S_t) \geq Ke^{-r(T-t)} - S_t, \quad 0 \leq t \leq T.$$

The results in this section are model-independent. To say more about option values we need a model for the dynamic evolution of a stock price. One of the simplest continuous-time models is the Black-Scholes-Merton (BSM) model, which we shall describe later, and one of the simplest discrete-time models is the binomial model, which we shall also see shortly.

1.5.2 Combinations of options

Options can be combined to give a variety of payoffs for different hedging purposes, or for speculation on movements in the underlying asset price, and they are often used to do so because the option premiums are relatively small in some cases, thus proving attractive to gamblers.

A *straddle* is a call and a put with the same strike and maturity. The payoff of a long position in a straddle is

$$(S_T - K)^+ + (K - S_T)^+ = \begin{cases} K - S_T, & S_T < K \\ S_T - K, & S_T \geq K, \end{cases} \tag{1.7}$$

This payoff is illustrated in Figure 5.

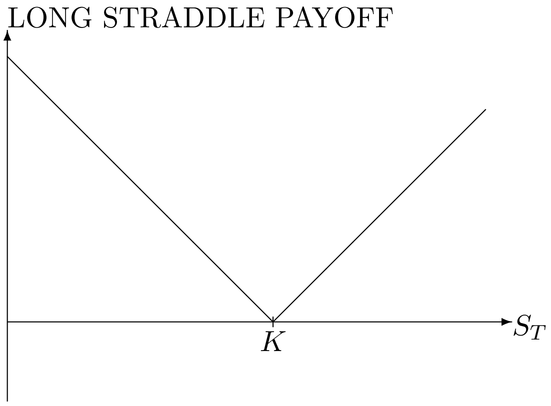


Figure 5: Straddle payoff

1.6 Some history*

As remarked earlier, the origins of derivatives lie in medieval agreements between farmers and merchants to insure farmers against low crop prices.

In the 1860s the Chicago Board of Trade was founded to trade commodity futures (contracts that set trading prices of commodities in advance), formalising the act of hedging against future price changes of important products.

Options were first valued by Bachelier in 1900 in his PhD thesis, a translation of which can be found in the book by Davis and Etheridge [3]. Bachelier introduced a stochastic process now known as *Brownian motion* (BM) to model stock price movements in continuous time. Bachelier did this before a rigorous treatment of BM was available in mathematics. His work was decades ahead of its time, both mathematically and economically speaking, and was therefore not given the credit it deserved at the time. In the decades that followed, mathematicians and physicists (Einstein, Wiener, Lévy, Kolmogorov, Feller to name but a few) developed a rigorous theory of Brownian motion, and Itô developed a rigorous theory of *stochastic integration* with respect to Brownian motion, leading to the notion of a *stochastic calculus*, which we shall encounter. In the 1960s, economists re-discovered Bachelier's work, and this was one of the ingredients that led to the modern theory of option valuation.

In the early 1970s a combination of forces existed which made markets more risky, derivatives more prominent, and their valuation and trading possible. The system of fixed exchange rates that existed before 1970 collapsed, and the Middle East oil crises caused a big increase in the volatility of financial prices. This increased the demand for risk management products such as options. At the same time Black and Scholes [2] and Merton [10] (BSM) published their seminal work on how to price options, based on managing the risk associated with selling such an asset. This breakthrough, for which Scholes and Merton received a Nobel

Prize (Black having passed away in 1995) coincided with the opening of the Chicago Board Options Exchange (CBOE), giving individuals both a means to value option contracts and a marketplace where they could profit from this knowledge of the fair price.

Following on from this, the financial deregulation of the 1980s, allied to technological developments which made it possible to trade securities globally and to run large portfolios of complex products, caused a huge increase in risky trading across international borders. This opened up yet more risks across currencies, interest rates and equities, and financial institutions very skillfully (or opportunistically, perhaps) created markets to trade derivatives and to sell these products to customers. This has led to the massive increase in derivative trading that we now see, with the volume of derivative contracts traded now dwarfing that in the associated underlying assets.

The papers of Black-Scholes [2] and Merton [10] attracted mathematicians to the subject, and led to a mathematically rigorous approach to valuing derivatives, based on probability and martingale theory, inspired by Harrison and Pliska [6]. This led directly to modern financial mathematics, and has also contributed to the advent of derivatives written on a plethora of underlying stochastic reference entities, such as interest rates, weather indices, default events, as well as on more traditional traded underlying securities such as stocks, currencies and interest rates.

Part II
The binomial model

2 Coin-toss space: a finite probability space

The binomial model is a simple dynamic model of a stock price process in which a fictitious coin is tossed, and a stock price depends on the outcome of the coin tosses.

Let $\mathbb{T} := \{0, 1, \ldots, n\}$ represent a discrete time set. Let $\Omega = \Omega_n$, the set of all outcomes of n coin tosses. The finite set Ω is called the *sample space*, with elements $\omega \in \Omega$ called *sample points*, representing the possible outcomes of the random experiment in which the coin is tossed. Each sample point $\omega \in \Omega$ is a sequence of length n, written as $\omega = \omega_1\omega_2\ldots\omega_n$, where each $\omega_t, t \in \mathbb{T}$ is either H (head) ot T (tail), representing the outcome of the t^{th} coin toss.

Let $\mathcal{F}$ be the set of all subsets of Ω; $\mathcal{F}$ is a *σ-algebra* (or *σ-field*), that is, a collection of subsets of Ω with the properties: (i) $\emptyset \in \mathcal{F}$, (ii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, (iii) if $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}$, then $\cup_{k=1}^{\infty} A_k$ is also in $\mathcal{F}$. We interpret σ-algebras as a record of information (as we shall see shortly). The pair $(\Omega, \mathcal{F})$ is called a *measurable space.*

We place a *probability measure* $\mathbb{P}$ on $(\Omega, \mathcal{F})$. A probability measure $\mathbb{P}$ is a function mapping $\mathbb{P} : \mathcal{F} \to [0, 1]$ with the properties: (i) $\mathbb{P}(\Omega) = 1$, (ii) if $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{F}$, then $\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$. The interpretation is that, for a set $A \in \mathcal{F}$, there is a probability in $[0, 1]$ that the outcome of a random experiment will lie in the set A. We think of $\mathbb{P}(A)$ as this probability. The set $A \in \mathcal{F}$ is called an *event.*

For $A \in \mathcal{F}$ we define

$$\mathbb{P}(A) := \sum_{\omega \in A} \mathbb{P}(\omega). \tag{2.1}$$

We can define $\mathbb{P}(A)$ in this way because A has only finitely many elements.

Suppose the probability of H on each coin toss is $p \in (0,1)$, then the probability of T is $q := 1-p$. For each $\omega = (\omega_1\omega_2\ldots\omega_n) \in \Omega$ we define

$$\mathbb{P}(\omega) := p^{\text{Number of H in }\omega} q^{\text{Number of T in }\omega}. \tag{2.2}$$

Then for each $A \in \mathcal{F}$ we define $\mathbb{P}(A)$ according to (2.1).

Definition 2.1 (Filtration)**.** A *filtration* $F = (\mathcal{F}_t)_{t\in\mathbb{T}}$ is a sequence of increasing σ-algebras $\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n$. That is, $\mathcal{F}_s \subset \mathcal{F}_t$ if $s < t$.

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$, with each $\mathcal{F}_t \subset \mathcal{F}$, is called a *filtered probability space.*

An $\mathbb{R}$-valued *random variable* $X \equiv X(\omega)$ on $(\Omega, \mathcal{F})$ is a measurable function $X : \Omega \to \mathbb{R}$, that is, the set $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \equiv \{X \in A\} \in \mathcal{F}$, for every Borel set $A \subset \mathbb{R}$.

Since a random variable maps Ω into $\mathbb{R}$, we can look at the preimage, under the random variable, of sets in $\mathbb{R}$, that is, sets of the form

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A \subset \mathbb{R}\} \equiv \{X \in A\},$$

which is, of course, a subset of Ω. The complete list of subsets of Ω that you can get as preimages (under X) of sets in $\mathbb{R}$, turns out to be a σ-algebra, whose content is exactly the information obtained by observing X, and is called the *σ-algebra generated by the random variable X.*

Definition 2.2. Let Ω be a nonempty finite set and let $\mathcal{F}$ be the σ-algebra of all subsets of Ω. Let X be a random variable on $(\Omega, \mathcal{F})$. The σ-algebra $\sigma(X)$ generated by X is defined to be the collection of all sets of the form $\{\omega \in \Omega : X(\omega) \in A\}$ where A is a subset of $\mathbb{R}$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. We say that X is *$\mathcal{G}$-measurable* if every set in $\sigma(X)$ is also in $\mathcal{G}$.

Definition 2.3 (Induced measure (distribution) of a random variable X)**.** Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. For $A \subseteq \mathbb{R}$, we define the *induced measure* of the set A to be

$$\mu_X(A) := \mathbb{P}\{\omega \in \Omega : X(\omega) \in A\} \equiv \mathbb{P}\{X \in A\}.$$

So the induced measure of a set A tells us the probability that X takes a value in A. By the *distribution* of a random variable X, we mean any of the several ways of characterizing μ_X (so we shall sometimes refer to the induced measure μ_X associated with a random variable X as the distribution of X).

Remark 2.4. We make a clear distinction between random variables and their distributions. A random variable is a mapping from Ω to $\mathbb{R}$, nothing more, and has an existence quite apart from any discussion of probabilities. The distribution of a random variable is a measure μ_X on $\mathbb{R}$, that is, a way of assigning probabilities to sets in $\mathbb{R}$. It depends on the random variable X *and* on the probability measure $\mathbb{P}$ we use on Ω. If we change $\mathbb{P}$, we change the distribution of the random variable X, but not the random variable itself. Thus, a random variable can have more than one distribution (e.g. an objective or "market" distribution, and a "risk-neutral" distribution, and we shall see such constructs in finance). In a similar vein, two *different random variables* can have the *same distribution.*

Definition 2.5. Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. The *expected value* or *expectation* of X is defined to be

$$\mathbb{E}[X] := \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega). \tag{2.3}$$

Notice that this is a sum over the *sample space* Ω. When Ω is finite and the random variable X takes on a finite number of values, this sum over Ω can be converted to a sum over $\mathbb{R}$, as shown in the Background Probability notes. It is easy to see that for an indicator function $\mathbb{1}_A$ of an event $A \in \mathcal{F}$, the definition of expectation leads to $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$.

Remark 2.6. When the sample space Ω is infinite and, in particular, uncountable, the summation in the definition of expectation is replaced by an integral. In general, the integral over an abstract measurable space $(\Omega, \mathcal{F})$ with respect to a probability measure $\mathbb{P}$ is a so-called *Lebesgue integral* (which has all the linearity and comparison properties we associate with ordinary integrals). The expectation $\mathbb{E}[X]$ becomes the Lebesgue integral over Ω of X with respect to $\mathbb{P}$, written as

$$\mathbb{E}[X] = \int_\Omega X \,\mathrm{d}\mathbb{P} \equiv \int_\Omega X(\omega) \,\mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} x \,\mathrm{d}\mu_X(x) \equiv \int_\mathbb{R} x\mu_X(\,\mathrm{d}x). \tag{2.4}$$

When X takes on a continuum of values and has a density f_X, then $\mathrm{d}\mu_X(x) = f_X(x)\,\mathrm{d}x$ and the integral on the right-hand side if (2.4) reduces to the familiar Riemann integral $\int_\mathbb{R} x f_X(x)\,\mathrm{d}x$.

We do not delve into the construction of Lebesgue integrals over abstract spaces here. Merely think of the RHS of (2.4) as an alternative notation for the sum $\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$. See Williams [14] or Shreve [13] for more details on Lebesgue integration.

Definition 2.7 (Discrete-time stochastic process)**.** let $\mathbb{T} = \{0\} \cup \mathbb{N} = \{0, 1, 2, \ldots\}$ be a discrete time set. A discrete-time *stochastic process* $(X_t)_{t\in\mathbb{T}}$ is a sequence of random variables.

Definition 2.8. Let $\mathbb{T} = \{0\} \cup \mathbb{N}$. A stochastic process $(X_t)_{t\in\mathbb{T}}$ on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}})$ is *adapted* to the filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ if X_t is $\mathcal{F}_t$-measurable for each $t \in \mathbb{T}$.

3 Conditional expectation

Definition 3.1 (Conditional expectation)**.** Let let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. The *conditional expectation* $\mathbb{E}[X|\mathcal{G}]$ is defined to be any random variable Y that satisfies

1. $Y = \mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable

2. For every set $A \in \mathcal{G}$, we have the *partial averaging property*

$$\int_A Y \,\mathrm{d}\mathbb{P} \equiv \int_A \mathbb{E}[X|\mathcal{G}] \,\mathrm{d}\mathbb{P} = \int_A X \,\mathrm{d}\mathbb{P}.$$

Note (we do not prove this here) that there is always a random variable Y satisfying the above properties (provided that $\mathbb{E}|X| < \infty$), i.e. conditional expectations always exist. There can be more than one random variable satisfying the above properties, but if Y' is another one, then $Y = Y'$ with probability 1 (or almost surely (a.s.)).

For random variables X, Y it is standard notation to write

$$\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)].$$

3.1 Partial averaging

The partial averaging property is

$$\int_A \mathbb{E}[X|\mathcal{G}]\,\mathrm{d}\mathbb{P} = \int_A X\,\mathrm{d}\mathbb{P}, \quad \forall A \in \mathcal{G}.$$

We can rewrite this as

$$\mathbb{E}[\mathbb{1}_A \cdot \mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[\mathbb{1}_A \cdot X]. \tag{3.1}$$

Note that $\mathbb{1}_A(\omega)$ (which equals 1 for $\omega \in A$ and 0 otherwise) is a $\mathcal{G}$-measurable random variable. Equation (3.1) suggests (and it is indeed true) that the following holds (see the Background Probability notes for a proof).

Lemma 3.2. *If V is any $\mathcal{G}$-measurable random variable, then provided $\mathbb{E}[V \cdot \mathbb{E}[X|\mathcal{G}]] < \infty$,*

$$\mathbb{E}[V \cdot \mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[V \cdot X]. \tag{3.2}$$

Based on Lemma 3.2, we can replace the second condition in the definition of conditional expectation by (3.2), so that the defining properties of $Y = \mathbb{E}[X|\mathcal{G}]$ are:

1. $Y = \mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable
2. For every $\mathcal{G}$-measurable random variable V, we have
$$\mathbb{E}[V \cdot \mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[V \cdot X].) \tag{3.3}$$

Notice that we can write (3.3) as

$$\mathbb{E}[V \cdot (\mathbb{E}[X|\mathcal{G}] - X)] = 0,$$

which allows an interpretation of $\mathbb{E}[X|\mathcal{G}]$ as the projection of the vector X on to the subspace $\mathcal{G}$. Then $\mathbb{E}[X|\mathcal{G}] - X$ is perpendicular to any V in the subspace.

3.2 Properties of conditional expectation

Proofs of the properties below are given in the Background Probability notes. All the X below satisfy $\mathbb{E}|X| < \infty$.

1. $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.

   The conditional expectation of X is thus an unbiased estimator of the random variable X.

2. If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$.

   In other words, if the information content of $\mathcal{G}$ is sufficient to determine X, then the best estimate of X based on $\mathcal{G}$ is X itself.

3. (Linearity) For $a_1, a_2 \in \mathbb{R}$,
$$\mathbb{E}[a_1 X_1 + a_2 X_2|\mathcal{G}] = a_1\mathbb{E}[X_1|\mathcal{G}] + a_2\mathbb{E}[X_2|\mathcal{G}].$$

4. (Positivity) If $X \geq 0$ almost surely, then $\mathbb{E}[X|\mathcal{G}] \geq 0$ almost surely.

5. (Jensen's inequality) If $\phi : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}|\phi(X)| < \infty$, then
$$\mathbb{E}[\phi(X)|\mathcal{G}] \geq \phi(\mathbb{E}[X|\mathcal{G}]).$$

6. (Tower property) If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then
$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}], \quad \text{a.s.}$$

   The intuition here is that $\mathcal{G}$ contains more information than $\mathcal{H}$. If we estimate X based on the information in $\mathcal{G}$, and then estimate the estimator based on the smaller amount of information in $\mathcal{H}$, then we get the same result as if we had estimated X directly based on the information in $\mathcal{H}$.

7. (Taking out what is known). If Z is $\mathcal{G}$-measurable, then
$$\mathbb{E}[ZX|\mathcal{G}] = Z \cdot \mathbb{E}[X|\mathcal{G}].$$

8. (Role of independence) If X is independent of $\mathcal{H}$ (i.e. if $\sigma(X)$ and $\mathcal{H}$ are independent σ-algebras), then
$$\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]. \tag{3.4}$$

   The intuition behind (3.4) is that if X is independent of $\mathcal{H}$, then the best estimate of X based on the information in $\mathcal{H}$ is $\mathbb{E}[X]$, the same as the best estimate of X based on no information.

4 Martingales

Definition 4.1 (Martingale)**.** A stochastic process $M = (M_t)_{t=0}^T$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t=0}^T, \mathbb{P})$ is a *martingale* with respect to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t=0}^T$ if: (i) each M_t is $\mathcal{F}_t$-measurable (so the process $(M_t)_{t=0}^T$ is dapted to the filtration $(\mathcal{F}_t)_{t=0}^T$); (ii) for each $t \in \{0, 1, \dots, T\}$, $\mathbb{E}[|M_t|] < \infty$, and (iii):
$$\mathbb{E}[M_{t+1}|\mathcal{F}_t] = M_t, \quad t = 0, 1, \dots, T-1.$$

So martingales tend to go neither up nor down. A *supermartingale* tends to go *down*, i.e. the second condition above is replaced by $\mathbb{E}[M_{t+1}|\mathcal{F}_t] \leq M_t$. A *submartingale* tends to go *up*, i.e. $\mathbb{E}[M_{t+1}|\mathcal{F}_t] \geq M_t$.

A simple argument using the tower property and induction shows the following.

Lemma 4.2.
$$\mathbb{E}[M_{t+u}|\mathcal{F}_t] = M_t,$$
for arbitrary $u \in \{1, \ldots, T-t\}$.

Proof. Consider $\mathbb{E}[M_{t+2}|\mathcal{F}_t]$. By the tower property,
$$\mathbb{E}[M_{t+2}|\mathcal{F}_t] = \mathbb{E}[\mathbb{E}[M_{t+2}|\mathcal{F}_{t+1}]|\mathcal{F}_t] = \mathbb{E}[M_{t+1}|\mathcal{F}_t] = M_t,$$
and continuing in this fashion we get
$$\mathbb{E}[M_{t+u}|\mathcal{F}_t] = M_t, \quad \text{for } u = 1, 2, \ldots, T-t.$$
□

Lemma 4.3. *Let* X *be an integrable random variable (*$\mathbb{E}[|X|] < \infty$*) on a filtered probability space* $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t=0}^T, \mathbb{P})$*. Define*
$$M_t := \mathbb{E}[X|\mathcal{F}_t], \quad t \in \{0, 1, \ldots, T\}.$$
Then $M := (M_t)_{t=0}^T$ *is a* $(\mathbb{P}, \mathbb{F})$*-martingale.*

Proof. We have
$$\begin{aligned} \mathbb{E}[M_{t+1}|\mathcal{F}_t] &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{t+1}]|\mathcal{F}_t] \\ &= \mathbb{E}[X|\mathcal{F}_t] \quad \text{(by the tower property)} \\ &= M_t. \end{aligned}$$
□

Definition 4.4 (Predictable process)**.** A a *predictable* process $(\xi_t)_{t=1}^T$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0}^T, \mathbb{P})$ s one such that, for each $t \in \{1, \ldots, T\}$, ξ_t is $\mathcal{F}_{t-1}$-measurable.

The following two propositions are proven in the Background Probability notes.

Proposition 4.5. *Let* $(M_t)_{t=0}^T$ *be a martingale on a filtered probability space* $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t=0}^T, \mathbb{P})$*. Let* $(\xi_t)_{t=1}^T$ *be a bounded predictable process. Then the process* $N := (N_t)_{t=0}^T$ *defined by*
$$N_0 := 0, \quad N_t := \sum_{s=1}^t \xi_s (M_s - M_{s-1}), \quad t = 1, \ldots, T,$$
is a $(\mathbb{P}, \mathbb{F})$*-martingale.*

Remark 4.6. The process N is called a *martingale transform* or discrete-time stochastic integral, and is sometimes denoted $N_t = \int_0^t \xi_s \, \mathrm{d}M_s$ or $N_t = (\xi \cdot M)_t$.

Proposition 4.7. *On a filtered probability space* $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t=0}^T, \mathbb{P})$*, an adapted sequence of real random variables* $M = (M_t)_{t=0}^T$ *is a* $(\mathbb{P}, \mathbb{F})$*-martingale if and only if for any predictable process* $\xi = (\xi_t)_{t=1}^T$*, we have*
$$\mathbb{E}\left[\sum_{s=1}^t \xi_s (M_s - M_{s-1})\right] = 0, \quad t = 1, \ldots, T.$$

5 Equivalent measures

Definition 5.1 (Absolutely continuous measures)**.** Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(\Omega, \mathcal{F})$. Assume that for every $A \in \mathcal{F}$ satisfying $\mathbb{P}(A) = 0$, we also have $\mathbb{Q}(A) = 0$. Then we say $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$, written $\mathbb{Q} \ll \mathbb{P}$.

Here is a deep theorem, which we do not prove.

Theorem 5.2 (Radon-Nikodym)**.** *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(\Omega, \mathcal{F})$, such that $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$. Under this assumption, there is a nonnegative random variable Z such that*

$$\mathbb{Q}(A) = \int_A Z \, \mathrm{d}\mathbb{P}, \quad \forall A \in \mathcal{F}, \tag{5.1}$$

and Z is called the Radon-Nikodym derivative *of $\mathbb{Q}$ with respect to $\mathbb{P}$.*

The random variable Z is often written as

$$Z = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}.$$

Equation (5.1) implies the apparently stronger condition

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}[XZ]$$

for every random variable X for which $\mathbb{E}[XZ] < \infty$. To see this, note that (5.1) in Theorem 5.2 is equivalent to

$$\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_A] = \mathbb{E}[\mathbb{1}_A Z], \quad A \in \mathcal{F}.$$

This is then extended this to general X via an argument that is called the "standard machine" in Section 1.5 of Shreve [13] (see the Background Probability notes for examples of this).

Definition 5.3 (Equivalent measures)**.** If $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$ and $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$, then we say that $\mathbb{P}$ and $\mathbb{Q}$ are *equivalent*, written $\mathbb{Q} \sim \mathbb{P}$.

In other words, $\mathbb{P}$ and $\mathbb{Q}$ are equivalent if and only if

$$\mathbb{P}(A) = 0 \quad \text{exactly when } \mathbb{Q}(A) = 0, \quad \forall A \in \mathcal{F}.$$

If $\mathbb{P}$ and $\mathbb{Q}$ are equivalent and Z is the Radon-Nikodym derivative of $\mathbb{Q}$ w.r.t. $\mathbb{P}$, then $\frac{1}{Z}$ is the Radon-Nikodym derivative of $\mathbb{P}$ w.r.t. $\mathbb{Q}$, i.e.

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}[XZ] \quad \forall X, \tag{5.2}$$

$$\mathbb{E}[Y] = \mathbb{E}^{\mathbb{Q}}\left[Y \cdot \frac{1}{Z}\right] \quad \forall Y, \tag{5.3}$$

and letting X and Y be related by $Y = XZ$ we see that the above two equations are the same.

Example 5.4 (Radon-Nikodym theorem in 2-period coin toss space). Let $\Omega = \Omega_2$ given by

$$\Omega_2 = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\},$$

the set of coin toss sequences of length 2. Let $\mathbb{P}$ correspond to probability $\frac{1}{3}$ for H and $\frac{2}{3}$ for T, and let $\mathbb{Q}$ correspond to probability $\frac{1}{2}$ for H and $\frac{1}{2}$ for T. Then the Radon-Nikodym derivative of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ is easily seen to be

$$Z(\omega) = \frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)}, \quad \omega \in \Omega,$$

so that

$$Z(\mathrm{HH}) = \frac{9}{4}, \quad Z(\mathrm{HT}) = \frac{9}{8}, \quad Z(\mathrm{TH}) = \frac{9}{8}, \quad Z(\mathrm{TT}) = \frac{9}{16}.$$

6 The binomial stock price process

On the n-period coin toss space Ω_n, with $\mathbb{T} = \{0, 1, \ldots, n\}$, for $t \in \mathbb{T}$, denote by $\mathcal{F}_t$ the σ-algebra generated by the first t coin tosses. Then, $\mathcal{F}_t$ is a collection of subsets $A \subset \Omega = \Omega_n$ such that $\mathcal{F}_t$ is a σ-algebra, and such that if one has the information of the results of the first t coin tosses (but is not told the outcome ω of all n coin tosses), then one can say whether $\omega \in A$ or $\omega \notin A$, for each $A \in \mathcal{F}_t$. Then $\mathbb{F} := (\mathcal{F}_t)_{t\in\mathbb{T}}$ is a filtration which records how information unfolds as one observes the results of successive coin tosses, as we will see in an example shortly.

The binomial model contains two assets, a *riskless asset* (or cash account, or money market account, or bond), with price process $S^0 = (S^0_t)_{t\in\mathbb{T}}$, and a *risky asset* or *stock*, with price process $S = (S_t)_{t\in\mathbb{T}}$.

The process S^0 evolves according to

$$S^0_{t+1} = (1+r)S^0_t, \quad t = 0, 1, \ldots, n-1, \quad S^0_0 = 1,$$

where $r \geq 0$ is the one-period (assumed constant) interest rate. Hence we have

$$S^0_t = (1+r)^t, \quad t = 0, 1, \ldots, n, \tag{6.1}$$

and S^0_t represents the value of at time t of one unit of currency (say \$1) invested at time zero.

Regarding the stock price process $S = (S_t)_{t\in\mathbb{T}}$, for each $t \in \mathbb{T}$, $S_t \equiv S_t(\omega)$ (for $\omega \in \Omega$) is a one-dimensional random variable on the measurable space $(\Omega, \mathcal{F})$, such that $S_t(\omega) = S_t(\omega_1 \ldots \omega_t)$ is the stock price after t coin tosses. The sequence of random variables $S = (S_t)_{t\in\mathbb{T}}$ is a stochastic process. We shall see that for each $t \in \mathbb{T}$, S_t is $\mathcal{F}_t$-measurable, so that S is an $\mathbb{F}$-adapted process. This encapsulates the idea that the information at time $t \in \mathbb{T}$, represented by $\mathcal{F}_t$, is sufficient information to know the values of S_s for all $s \leq t$.

Define two constants u, d satisfying $u > 1 + r > d > 0$. The evolution of the stock price is given by (see Figure 6)

$$S_{t+1}(\omega) = \begin{cases} S_t u, & \text{if } \omega_{t+1} = \mathrm{H}, \\ S_t d, & \text{if } \omega_{t+1} = \mathrm{T}, \end{cases} \qquad t = 0, 1, \ldots, n-1.$$

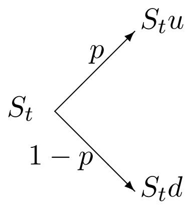


Figure 6: Binomial process for stock price. We have associated a probability $p \in (0,1)$ with an upward stock price move.

We also write

$$S_{t+1}(\omega_1 \dots \omega_{t+1}) = \begin{cases} S_t(\omega_1 \dots \omega_t)u, & \text{if } \omega_{t+1} = \mathrm{H} \\ S_t(\omega_1 \dots \omega_t)d, & \text{if } \omega_{t+1} = \mathrm{T}, \end{cases} \qquad t = 0, 1, \dots, n-1,$$

whenever we wish to emphasise that S_t actually depends only on the outcome of the first t coin tosses, and we abbreviate this notation further by sometimes suppressing the depndence on $\omega_1 \dots \omega_t$, and writing

$$S_{t+1}(\omega_{t+1}) = \begin{cases} S_t u, & \text{if } \omega_{t+1} = \mathrm{H} \\ S_t d, & \text{if } \omega_{t+1} = \mathrm{T}, \end{cases} \qquad t = 0, 1, \dots, n-1,$$

At time $t \in \mathbb{T}$ the possible stock prices are $S_t^{(j)}$, given by

$$S_t^{(j)} = S_0 u^j d^{t-j}, \quad j = 0, 1, \dots, t, \quad t \in \mathbb{T}.$$

Example 6.1 (One-period binomial model). Let $n = 1$, so $\mathbb{T} = \{0, 1\}$, and $\Omega = \Omega_1$ is the finite set

$$\Omega_1 := \{\mathrm{H}, \mathrm{T}\},$$

the set of outcomes of a single coin toss. The stock price process is $(S_t)_{t \in \{0,1\}}$, and $S_1(\omega)$ takes on two possible values, $S_1(\mathrm{H})$ or $S_1(\mathrm{T})$, given by

$$S_1(\omega) = \begin{cases} S_0 u, & \text{if } \omega = \mathrm{H} \\ S_0 d, & \text{if } \omega = \mathrm{T}. \end{cases}$$

6.1 Information and conditional expectation in the binomial model

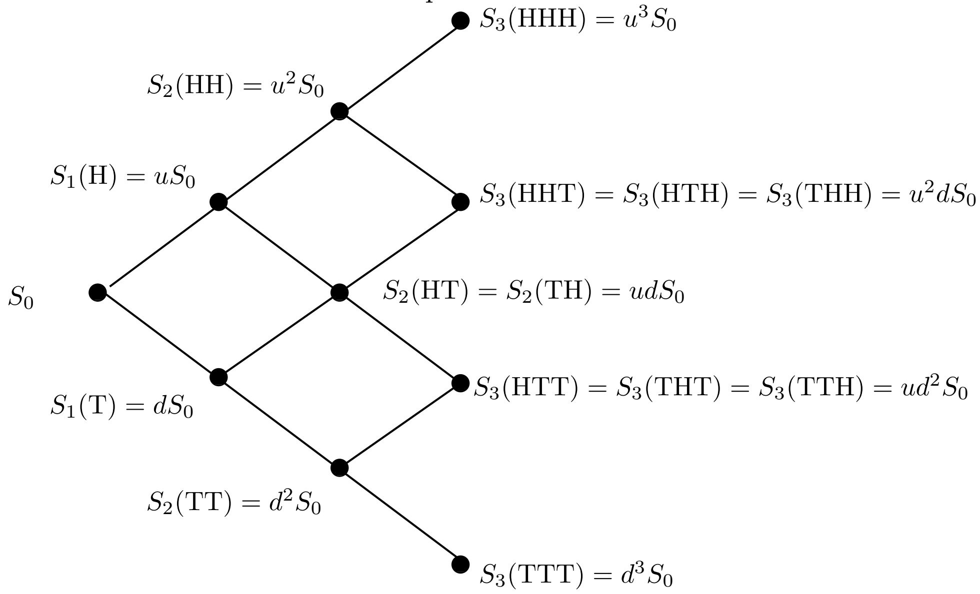
A filtration $\mathbb{F} = (\mathcal{F})_{t \in \mathbb{T}}$ captures information flow as time marches forward. We illustrate this, and the associated idea that the binomial stock price process $(S_t)_{t \in \mathbb{T}}$ is adapted to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ generated by the coin tosses,, along with some examples of conditional expectation, in a 3-period model.

Example 6.2 (3-period binomial model). Let $n = 3$, so that $\Omega = \Omega_3$, given by the finite set

$$\Omega = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}, \mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT}\},$$

the set of all possible outcomes of three coin tosses.

We can write down all the stock prices in a binomial tree as:



Define the following two subsets of Ω:

$$A_{\mathrm{H}} = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}\}, \quad A_{\mathrm{T}} = \{\mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT}\},$$

corresponding to the events that the first coin toss results in H and T respectively.

Let the coin have probability $p \in (0, 1)$ for H and $q := 1 - p$ for T. Using (2.2) the definition $\mathbb{P}(A) := \sum_{\omega \in A} \mathbb{P}(\omega)$ we find

$$\begin{aligned} \mathbb{P}(A_{\mathrm{H}}) &= \mathbb{P}\{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}\} = \mathbb{P}\{\text{H on first toss}\} = p, \\ \mathbb{P}(A_{\mathrm{T}}) &= \mathbb{P}\{\mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT}\} = \mathbb{P}\{\text{T on first toss}\} = q, \end{aligned}$$

precisely in accordance with intuition.

Here are two σ-algebras of subsets of Ω.

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \quad \mathcal{F}_1 = \{\emptyset, \Omega, A_H, A_T\}.$$

It is easy to see (exercise, or see Problem Sheet 0) that $\mathcal{F}_0, \mathcal{F}_1$ are both σ-algebras.

The σ-algebra $\mathcal{F}_1$ contains the “information of the first toss” or the “information up to time 1”. If one has infornation on the first toss only, then one cannot say what the actual outcome $\omega = \omega_1\omega_2\omega_3$ of all three coin tosses is. With information up to time 1, all one knows is that either $\omega_1 = \mathrm{H}$ or that $\omega_1 = \mathrm{T}$. In this case one can answer the question “is $\omega \in A$?” for every set in $\mathcal{F}_1$. One cannot answer a question such as: “is $\omega \in \{\mathrm{HHH}, \mathrm{HHT}\}$?” (one would need to know the outcome of the first two tosses to answer such a question). This is why $\mathcal{F}_1 = \{\emptyset, A_{\mathrm{H}}, A_{\mathrm{T}}, \Omega\}$.

The general principle is:

Fact 6.3. The σ-algebra $\mathcal{F}_t$ corresponding to information at time $t \in \mathbb{T}$ is composed of all the sets A such that $\mathcal{F}_t$ is indeed a σ-algebra, and such that one can answer the question: “is $\omega \in A$?”, given that one has information on the outcome of the first t coin tosses.

The trivial σ-algebra $\mathcal{F}_0$ contains no information: knowing whether the outcome ω of the three tosses is in $\emptyset$ (it is not) and whether it is in Ω (it is) tells you nothing about ω, in accordance with the idea that at time zero one know nothing about the eventual outcome ω of the three coin tosses. All one can say is that $\omega \notin \emptyset$ and $\omega \in \Omega$, and so $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Define

$$\begin{aligned} A_{\mathrm{HH}} &= \{\mathrm{HHH}, \mathrm{HHT}\}, & A_{\mathrm{HT}} &= \{\mathrm{HTH}, \mathrm{HTT}\}, \\ A_{\mathrm{TH}} &= \{\mathrm{THH}, \mathrm{THT}\}, & A_{\mathrm{TT}} &= \{\mathrm{TTH}, \mathrm{TTT}\}, \end{aligned}$$

corresponding to the events that the first two coin tosses result in HH, HT, TH and TT respectively. Consider the collection of sets

$$\mathcal{F}_2 = \{\emptyset, \Omega, A_{\mathrm{HH}}, A_{\mathrm{HT}}, A_{\mathrm{TH}}, A_{\mathrm{TT}}, \text{plus all unions of these}\}.$$

Then $\mathcal{F}_2$ can be written as follows (this is confirmed in Problem Sheet 0):

$$\begin{aligned} \mathcal{F}_2 = \; & \{\emptyset, \Omega, A_{\mathrm{HH}}, A_{\mathrm{HT}}, A_{\mathrm{TH}}, A_{\mathrm{TT}}, A_{\mathrm{H}}, A_{\mathrm{T}}, \\ & A_{\mathrm{HH}} \cup A_{\mathrm{TH}}, A_{\mathrm{HH}} \cup A_{\mathrm{TT}}, A_{\mathrm{HT}} \cup A_{\mathrm{TH}}, A_{\mathrm{HT}} \cup A_{\mathrm{TT}}, \\ & A^c_{\mathrm{HH}}, A^c_{\mathrm{HT}}, A^c_{\mathrm{TH}}, A^c_{\mathrm{TT}}\}. \end{aligned}$$

Then $\mathcal{F}_2$ is indeed a σ-algebra (a tedious and lengthy verification) which contains the "information of the first two tosses" or the "information up to time 2". This is because, if you know the outcome of the first two coin tosses, you can say whether the outcome $\omega \in \Omega$ of all three tosses satisfies $\omega \in A$ for each $A \in \mathcal{F}_2$.

Similarly, $\mathcal{F}_3 \equiv \mathcal{F}$, the set of all subsets of Ω, contains "full information" about the outcome of all three tosses. The sequence of σ-algebras $\mathbb{F} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\}$ is a filtration.

The stock price process $(S_t)_{t \in \mathbb{T}}$ is $\mathbb{F}$-adapted. That is, the value of the random variable S_t is known after t coin tosses, equivalently, S_t is $\mathcal{F}_t$-measurable, for each $t \in \mathbb{T}$, meaning that the event

$$\{\omega \in \Omega : S_t(\omega) \in A \subseteq \mathbb{R}\} = \{S_t \in A\},$$

is in $\mathcal{F}_t$.

At time zero, with $\mathcal{F}_0 = \{\Omega, \emptyset\}$, we must have that S_0 is a deterministic constant,

$$S_0(\omega) = a \in \mathbb{R}, \quad \forall \omega \in \Omega,$$

so that sets of the form $\{S_0 \in A \subseteq \mathbb{R}\}$ are clearly either in Ω or $\emptyset$, so S_0 is $\mathcal{F}_0$-measurable.

The random variable S_1 must be of the form

$$S_1(\omega) = \begin{cases} a \in \mathbb{R}, & \text{if } \omega \in A_{\mathrm{H}}, \\ b \in \mathbb{R}, & \text{if } \omega \in A_{\mathrm{T}}, \end{cases}$$

since the only information available at time 1 is whether $\omega_1 = \mathrm{H}$ or $\omega_1 = \mathrm{T}$, and we notice that S_1 is indeed of the above form.

Continuing to argue in this fashion, we find that at each time $t \in \mathbb{T}$, the event

$$\{\omega \in \Omega : S_t(\omega) \in A \subseteq \mathbb{R}\} = \{S_t \in A\},$$

is in $\mathcal{F}_t$. The stochastic process $S = (S_t)_{t \in \mathbb{T}}$ is said to be *adapted* to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$, as in Definition 2.8.

Now suppose $S_0 = 4$, $u = 2$ and $d = \frac{1}{2}$. Then S_0, S_1, S_2 and S_3 are all random variables, denoted by $S_t(\omega)$ for $t = 0, 1, 2, 3$ and $\omega \in \Omega$. We may calculate the values of $S_2(\omega)$ for all $\omega \in \Omega$, as

$$\begin{aligned} S_2(\mathrm{HHH}) &= S_2(\mathrm{HHT}) = 16, \\ S_2(\mathrm{HTH}) &= S_2(\mathrm{HTT}) = S_2(\mathrm{THH}) = S_2(\mathrm{THT}) = 4, \\ S_2(\mathrm{TTH}) &= S_2(\mathrm{TTT}) = 1. \end{aligned}$$

Now consider the preimage under the random variable S_2 of certain sets in $\mathbb{R}$. Specifically, consider the interval $[4, 29]$. The preimage under S_2 of this interval is

$$\{\omega \in \Omega : S_2(\omega) \in [4, 29]\}.$$

We can characterise the above subset of Ω in terms of one of the sets given earlier in the list of sets in $\mathcal{F}_2$. We have: $\{\omega \in \Omega : S_2(\omega) \in [4, 29]\} = A^c_{\mathrm{TT}}$.

Suppose we list, in as minimal a fashion as possible, the subsets of Ω that we can get as preimages under S_2 of sets in $\mathbb{R}$, along with sets which can be built by taking unions of these, then this collection of sets turns out to be a σ-algebra, the *σ-algebra generated by the random variable* S_2, denoted $\sigma(S_2)$.

Now, if $\omega \in A_{\mathrm{HH}}$, then $S_2(\omega) = 16$. If $\omega \in A_{\mathrm{HT}} \cup A_{\mathrm{TH}}$, then $S_2(\omega) = 4$. If $\omega \in A_{\mathrm{TT}}$, then $S_2(\omega) = 1$. Hence $\sigma(S_2)$ is composed of $\{\emptyset, \Omega, A_{\mathrm{HH}}, A_{\mathrm{HT}} \cup A_{\mathrm{TH}}, A_{\mathrm{TT}}\}$, plus all relevant unions and complements. Using the identities

$$\begin{aligned} A_{\mathrm{HH}} \cup (A_{\mathrm{HT}} \cup A_{\mathrm{TH}}) &= A^c_{\mathrm{TT}}, \\ A_{\mathrm{HH}} \cup A_{\mathrm{TT}} &= (A_{\mathrm{HT}} \cup A_{\mathrm{TH}})^c, \\ (A_{\mathrm{HT}} \cup A_{\mathrm{TH}}) \cup A_{\mathrm{TT}} &= A^c_{\mathrm{HH}}, \end{aligned}$$

we obtain

$$\sigma(S_2) = \{\emptyset, \Omega, A_{\mathrm{HH}}, A_{\mathrm{HT}} \cup A_{\mathrm{TH}}, A_{\mathrm{TT}}, A_{\mathrm{HH}} \cup A_{\mathrm{TT}}, A^c_{\mathrm{HH}}, A^c_{\mathrm{TT}}\}. \tag{6.2}$$

The information content of the σ-algebra $\sigma(S_2)$ is exactly the information learned by observing S_2. So, suppose the coin is tossed three times and you do not know the outcome ω, but you are told, for each set in $\sigma(S_2)$, whether ω is in the set. For instance, you might be told that ω is not in A_{HH}, is in $A_{\mathrm{HT}} \cup A_{\mathrm{TH}}$, and is not in A_{TT}. Then you know that in the first two throws there was a head and a tail, but you are not told in which order they occurred. This is the same information you would have got by being told that the value of $S_2(\omega)$ is 4.

Note that $\mathcal{F}_2$ contains all the sets which are in $\sigma(S_2)$, and even more. In other words, the information in the first two throws is greater than the information in S_2. In particular, if you see the first two tosses you can distinguish A_{HT} from A_{TH}, but you cannot make this distinction from knowing the value of S_2 alone.

Let us give an example of a conditional expectation: suppose we wish to estimate S_1, given S_2, and denote this estimate by $\mathbb{E}[S_1|S_2]$. This should have two properties: (i) it should be a random variable, so should depend on ω, $\mathbb{E}[S_1|S_2] = \mathbb{E}[S_1|S_2(\omega)] = \mathbb{E}[S_1|S_2](\omega)$, and (ii) it should be $\sigma(S_2)$-measurable, that is, if f the value of S_2 is known then the value of $\mathbb{E}[S_1|S_2]$ should also be known.

In particular, if $\omega = \text{HHH}$ or $\omega = \text{HHT}$, then $S_2(\omega) = u^2 S_0$ and, even without knowing ω, we know that $S_1(\omega) = uS_0$. We define

$$\mathbb{E}[S_1|S_2](\text{HHH}) := \mathbb{E}[S_1|S_2](\text{HHT}) = uS_0.$$

In other words

$$\mathbb{E}[S_1|S_2](\omega) = uS_0, \quad \omega \in A_{\text{HH}}.$$

Similarly we define

$$\mathbb{E}[S_1|S_2](\text{TTT}) := \mathbb{E}[S_1|S_2](\text{TTH}) = dS_0.$$

In other words

$$\mathbb{E}[S_1|S_2](\omega) = dS_0, \quad \omega \in A_{\text{TT}}.$$

Finally, if $\omega \in A = A_{\text{HT}} \cup A_{\text{TH}} = \{\text{HTH}, \text{HTT}, \text{THH}, \text{THT}\}$, then $S_2(\omega) = udS_0$, so that $S_1(\omega) = uS_0$ or $S_1(\omega) = dS_0$. So, to get $E[S_1|S_2]$ in this case, we take a weighted average, as follows. For $\omega \in A$ we define

$$\mathbb{E}[S_1|S_2](\omega) := \frac{\int_A S_1 \,\mathrm{d}\mathbb{P}}{\mathbb{P}(A)},$$

which is a partial average of S_1 over the set A, normalised by the probability of A.

Now, $\mathbb{P}(A) = 2pq$ and $\int_A S_1 \,\mathrm{d}\mathbb{P} = pq(u+d)S_0$, so that for $\omega \in A$

$$\mathbb{E}[S_1|S_2](\omega) = \frac{1}{2}(u+d)S_0, \quad \omega \in A = A_{\text{HT}} \cup A_{\text{TH}}.$$

(In other words, the best estimate of S_1, given that $S_2 = udS_0$, is the average of the possibilities uS_0 and dS_0.) Then we have that

$$\int_A \mathbb{E}[S_1|S_2] \,\mathrm{d}\mathbb{P} = \int_A S_1 \,\mathrm{d}\mathbb{P}.$$

In conclusion, we can write

$$\mathbb{E}[S_1|S_2](\omega) = g(S_2(\omega)),$$

where

$$g(x) = \begin{cases} uS_0, & \text{if } x = u^2 S_0 \\ \frac{1}{2}(u+d)S_0, & \text{if } x = udS_0 \\ dS_0, & \text{if } x = d^2 S_0. \end{cases}$$

In other words $\mathbb{E}(S_1|S_2)$ is random *only through dependence on* S_2 (and hence is $\sigma(S_2)$-measurable). This random variable satisfies

1. $\mathbb{E}[S_1|S_2]$ is $\sigma(S_2)$-measurable

2. For every $A \in \sigma(S_2)$,

$$\int_A \mathbb{E}[S_1|S_2] \,\mathrm{d}\mathbb{P} = \int_A S_1 \,\mathrm{d}\mathbb{P},$$

   which is the partial averaging property.

Here is another (simpler) example of conditional expectation in the 3-period binomial model. Recall the σ-algebra determined by the first toss, $\mathcal{F}_1 = \{\emptyset, \Omega, A_{\mathrm{H}}, A_{\mathrm{T}}\}$, where A_{H} (respectively A_{T}) is the event corresponding to a H (respectively a T) on the first toss.

Using the partial averaging property on the sets A_{H} and A_{T}, we can show (the obvious fact) that

$$\mathbb{E}[S_2|\mathcal{F}_1](\omega) = (pu + qd)S_1(\omega),$$

as follows: $\mathbb{E}[S_2|\mathcal{F}_1]$ is constant on A_{H} and on A_{T} (since it is $\mathcal{F}_1$-measurable) and must satisfy the partial averaging property on these sets:

$$\int_{A_{\mathrm{H}}} \mathbb{E}[S_2|\mathcal{F}_1]\,\mathrm{d}\mathbb{P} = \int_{A_{\mathrm{H}}} S_2\,\mathrm{d}\mathbb{P}, \quad \int_{A_{\mathrm{T}}} \mathbb{E}[S_2|\mathcal{F}_1]\,\mathrm{d}\mathbb{P} = \int_{A_{\mathrm{T}}} S_2\,\mathrm{d}\mathbb{P}.$$

(Obviously the partial averaging property is true on $\emptyset$ (all are zero) and it will be true on Ω if it is true on A_{H} and A_{T} since $A_{\mathrm{H}} \cup A_{\mathrm{T}} = \Omega$). On A_{H} we have

$$\begin{aligned}\int_{A_{\mathrm{H}}} \mathbb{E}[S_2|\mathcal{F}_1]\,\mathrm{d}\mathbb{P} &= \mathbb{P}(A_{\mathrm{H}})\mathbb{E}[S_2|\mathcal{F}_1](\omega) \quad (\text{since } \mathbb{E}[S_2|\mathcal{F}_1] \text{ is constant over } A_{\mathrm{H}}) \\ &= p\mathbb{E}[S_2|\mathcal{F}_1], \quad \forall \omega \in A_{\mathrm{H}},\end{aligned}$$

whilst on the other hand

$$\int_{A_{\mathrm{H}}} S_2\,\mathrm{d}\mathbb{P} = p^2u^2S_0 + pqudS_0.$$

Hence

$$\mathbb{E}[S_2|\mathcal{F}_1](\omega) = pu^2S_0 + qudS_0 = (pu + qd)uS_0 = (pu + qd)S_1(\omega), \quad \forall \omega \in A_{\mathrm{H}}.$$

Similarly, we can show that

$$\mathbb{E}[S_2|\mathcal{F}_1](\omega) = (pu + qd)S_1(\omega), \quad \forall \omega \in A_{\mathrm{T}}.$$

So overall we get

$$\mathbb{E}[S_2|\mathcal{F}_1](\omega) = (pu + qd)S_1(\omega), \quad \forall \omega \in \Omega.$$

With $\mathcal{F}_0 = \{\emptyset, \Omega\}$, we can show similarly that

$$\mathbb{E}[S_1|\mathcal{F}_0] = (pu + qd)S_0.$$

$\mathcal{F}_0$ contains no information, so any $\mathcal{F}_0$-measurable random variable must be constant (non-random). Therefore $\mathbb{E}[S_1|\mathcal{F}_0]$ is that constant which satisfies the averaging property

$$\int_{\Omega} \mathbb{E}[S_1|\mathcal{F}_0]\,\mathrm{d}\mathbb{P} = \int_{\Omega} S_1\,\mathrm{d}\mathbb{P} = \mathbb{E}[S_1] = (pu + qd)S_0,$$

and so we have

$$\mathbb{E}[S_1|\mathcal{F}_0] = (pu + qd)S_0.$$

We can generalise the above results to an n-period model.

Lemma 6.4. *In an n-period binomial model, we have*

$$\mathbb{E}[S_{t+1}|\mathcal{F}_t] = (pu + qd)S_t, \quad t = 0, 1, \ldots, n-1. \tag{6.3}$$

Proof. To show this, define, for any $t = 0, 1, \ldots, n-1$, the random variable

$$X := \frac{S_{t+1}}{S_t}.$$

Then $X = u$ if $\omega_{t+1} = \mathrm{H}$ and $X = d$ if $\omega_{t+1} = \mathrm{T}$, and X is independent of $\mathcal{F}_t$ because each coin toss is independent. Hence

$$\mathbb{E}[S_{t+1}|\mathcal{F}_t] = \mathbb{E}[XS_t|\mathcal{F}_t] = S_t\mathbb{E}[X|\mathcal{F}_t] = S_t\mathbb{E}[X] = (pu + qd)S_t.$$

□

Notice that the right-hand-side of (6.3) depends only on the current stock price S_t, signifying that the stock price process is a Markov process.

7 Derivative valuation in the binomial model

Take the standard n-period binomial stock price process S of Section 6 on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\in\mathbb{T}}, \mathbb{P})$, generated by n coin tosses, with time index set $\mathbb{T} = \{0, 1, 2, \ldots n\}$. The sample space Ω is finite, the probability measure $\mathbb{P}$ is called the *physical measure* (or objective measure, or the market measure). We assume $\mathcal{F}_n = \mathcal{F}$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

The sample space Ω is the set of all outcomes of n coin tosses, so each $\omega \in \Omega$ is of the form $\omega = (\omega_1\omega_2\ldots\omega_n)$, with each $\omega_t \in \{\mathrm{H}, \mathrm{T}\}$, for each $t \in \{1, \ldots, n\}$. The evolution of the stock price process $S = (S_t)_{t=0}^n$ is given by

$$S_{t+1} = \begin{cases} S_t u, & \text{if } \omega_{t+1} = \mathrm{H}, \\ S_t d, & \text{if } \omega_{t+1} = \mathrm{T}, \end{cases} \qquad t = 0, 1, \ldots, n-1,$$

where $u > 1 + r > d > 0$.

Introduce a financial agent with initial wealth X_0 at time zero, who can choose at each time how to split his wealth between the riskless and riskyt assets. The agent's trading strategy is the two-dimensional stochastic process

$$(\pi_t^0, \pi_t), \quad t \in \{1, \ldots, n\},$$

where, for $t \in \{1, \ldots, n\}$, π_t^0 denotes the number of units of the riskless asset held over the interval $[t-1, t)$ and π_t, denotes the number of units of the stock held over the interval $[t-1, t)$. The positions in the portfolio at time t, for $t \in \{1, \ldots, n\}$, are decided at time $t-1$ and kept until time t, when new asset price quotations are available.

Assumption 7.1. The portfolio process $(\pi_t^0, \pi_t)_{t\in\{1,\ldots,n\}}$ is *predictable*, so that for each $t \in \{1, \ldots, n\}$, π_t^0 and π_t are $\mathcal{F}_{t-1}$-measurable.

The initial wealth is given by

$$X_0 = \pi_1^0 + \pi_1 S_0. \quad \text{(budget constraint)} \tag{7.1}$$

Equation (7.1) is a *budget constraint*, that the agent splits all his initial wealth between cash and the risky stock.

The time 1 wealth is

$$X_1 = (1+r)\pi_1^0 + \pi_1 S_1, \tag{7.2}$$

where we have assumed that no wealth has been taken out of the portfolio for (say) consumption and no outside income has been injected into the portfolio. Equation (7.2) is thus one form of a *self-financing condition* on the portfolio wealth evolution. Using the budget constraint (7.1) we re-cast (7.2) into the form

$$X_1 = (1+r)X_0 + \pi_1(S_1 - (1+r)S_0). \tag{7.3}$$

Similar self-financing portfolio rebalancing occurs at each time $t \in \{1, \dots, n-1\}$. Define the wealth process $X = (X_t)_{t\in\mathbb{T}}$, where X_t denotes the wealth at time t (that is, at the end of the interval $[t-1, t)$ and the beginning of the interval $[t, t+1)$), for $t = 0, 1, \dots, n-1$, with X_n the final wealth at the end of the interval $[n-1, n]$. We then have the following evolution.

At the beginning of the interval $[t-1, t)$, and just after portfolio rebalancing has taken place, the wealth is X_{t-1} given by

$$X_{t-1} = \pi_t^0 S_{t-1}^0 + \pi_t S_{t-1} = \pi_t^0 (1+r)^{t-1} + \pi_t S_{t-1}, \tag{7.4}$$

where the last equality follows from the expression (6.1) for the value of the riskless asset at any time in $\mathbb{T}$. The position (π_t^0, π_t) is held over $[t-1, t)$, and the wealth X_t achieved at the end of this interval (and hence at the start of the interval $[t, t+1)$) is

$$X_t = \pi_t^0 S_t^0 + \pi_t S_t = \pi_t^0 (1+r)^t + \pi_t S_t. \tag{7.5}$$

At this time, t, the portfolio is rebalanced to (π_{t+1}^0, π_{t+1}) so that X_t is also given by

$$X_t = \pi_{t+1}^0 S_t^0 + \pi_{t+1} S_t.$$

Hence the general self-financing condition is

$$\pi_{t+1}^0 S_t^0 + \pi_{t+1} S_t = \pi_t^0 S_t^0 + \pi_t S_t, \quad t = 1, \dots, n-1.$$

We can raise this to a definition.

Definition 7.2. A trading strategy $(\pi_t^0, \pi_t)_{t=1}^n$ is *self-financing* if for every $t = 1, \dots, n-1$, we have

$$\pi_{t+1}^0 S_t^0 + \pi_{t+1} S_t = \pi_t^0 S_t^0 + \pi_t S_t, \quad t = 1, \dots, n-1.$$

Using (7.4) to eliminate π_t^0 from (7.5) we can write the portfolio wealth evolution as

$$X_t = (1+r)X_{t-1} + \pi_t(S_t - (1+r)S_{t-1}), \quad t = 1, \dots, n. \tag{7.6}$$

This can be put into a neater form if we work with *discounted* quantities, that is, we evaluate all quantities in units of the bond price. The discounted stock price process $\widetilde{S}$ is defined by

$$\widetilde{S}_t = \frac{S_t}{S_t^0} = \frac{S_t}{(1+r)^t}, \quad t = 0, 1, \ldots, n,$$

and the discounted wealth process is similarly defined by

$$\widetilde{X}_t = \frac{X_t}{S_t^0} = \frac{X_t}{(1+r)^t}, \quad t = 0, 1, \ldots, n.$$

Then, in terms of discounted quantities, the wealth evolution equation (7.6) becomes

$$\widetilde{X}_t = \widetilde{X}_{t-1} + \pi_t(\widetilde{S}_t - \widetilde{S}_{t-1}), \quad t = 1, \ldots, n. \tag{7.7}$$

Iterating this evolution from time zero to $t \in \mathbb{T}$ we obtain

$$\widetilde{X}_t = X_0 + \sum_{s=1}^{t} \pi_s(\widetilde{S}_s - \widetilde{S}_{s-1}), \quad t = 1 \ldots, n. \tag{7.8}$$

From this we see that the wealth process is completely specified by the initial wealth X_0 and the choice of stock portfolio π. When we need to emphasise the dependence of wealth on the chosen portfolio we write $X(\pi) \equiv X$.

The sum in (7.8) as the (discrete-time) *stochastic integral* of π with respect to $\widetilde{S}$, denoted by $(\pi \cdot \widetilde{S})$:

$$(\pi \cdot \widetilde{S})_t := \sum_{s=1}^{t} \pi_s(\widetilde{S}_s - \widetilde{S}_{s-1}), \quad t = 1, \ldots, n.$$

7.1 Equivalent martingale measures and no arbitrage

Definition 7.3 (Equivalent martingale measure)**.** An *equivalent martingale measure* (EMM), also called a *risk-neutral measure*, is a probability measure $\mathbb{Q} \sim \mathbb{P}$ such that the discounted stock price $\widetilde{S}$ is a $\mathbb{Q}$-martingale.

Lemma 7.4. *If a martingale measure exists, then the discounted wealth process of a self-financing portfolio process π is a $\mathbb{Q}$-martingale.*

Proof. If a martingale measure $\mathbb{Q}$ exists, we have $\mathbb{E}^{\mathbb{Q}}[S_t|\mathcal{F}_{t-1}] = S_{t-1}$. Then, from (7.7) we obtain

$$\mathbb{E}^{\mathbb{Q}}[\widetilde{X}_t|\mathcal{F}_{t-1}] = \widetilde{X}_{t-1}, \quad t = 1, \ldots, n,$$

so that the discounted wealth process is also a $\mathbb{Q}$-martingale.

$\square$

Remark 7.5. Lemma 7.4 also follows from the fact that the discounted wealth process is a finite sum of stochastic integrals, and hence a martingale transform (recall Proposition 4.5). For any self-financing strategy π, the discounted wealth process is given by (7.8), and combining this with Proposition 4.5, the discounted wealth process $\widetilde{X}$ is a $\mathbb{Q}$-martingale.

7.1.1 The risk-neutral measure in the binomial model

In the binomial model, there is a unique EMM $\mathbb{Q}$ defined as follows. For $t \in \{1, \ldots, n\}$, define

$$\mathbb{Q}(\omega_t = \mathrm{H}) = p^{\mathbb{Q}} := \frac{1+r-d}{u-d}, \quad \mathbb{Q}(\omega_t = \mathrm{T}) = q^{\mathbb{Q}} := \frac{u-(1+r)}{u-d}. \tag{7.9}$$

It is clear that $\mathbb{Q} \sim \mathbb{P}$, and that $\mathbb{E}^{\mathbb{Q}}[S_{t+1}|\mathcal{F}_t] = (1+r)S_t$, so that $\mathbb{Q}$ is indeed an EMM. We will see shortly that $\mathbb{Q}$ emerges naturally when we try to value derivatives in the binomial model via a replication argument. This is one manifestation of deep results called the *Fundamental Theorems of Asset Pricing* which hold in great generality (but are a little off-syllabus). We mention these theorems in passing.

7.1.2 Fundamental theorems of asset pricing

Recall the definition of arbitrage.

Definition 7.6 (Arbitrage)**.** An *arbitrage* over $\mathbb{T} = \{0, 1, \ldots, n\}$ is a strategy π such that the associated wealth process satisfies $X_0 = 0$, $\mathbb{P}[X_n(\pi) \geq 0] = 1$ and $\mathbb{P}[X_n(\pi) > 0] > 0$.

Theorem 7.7 (First Fundamental Theorem of Asset Pricing (FTAP I))**.** *A finite sample space, discrete-time financial market is arbitrage-free if and only if there exists an equivalent martingale measure.*

Remark 7.8. The proof of Theorem 7.7 is easy in one direction: suppose there exists an equivalent martingale measure $\mathbb{Q}$. Then for any self-financing strategy π we have, from Lemma 7.4, that the discounted wealth process is a $\mathbb{Q}$-martingale, so $\mathbb{E}^{\mathbb{Q}}[\widetilde{X}_n] = \widetilde{X}_0$. This immediately precludes the possibility of arbitrage. For suppose $\widetilde{X}$ is such that $\widetilde{X}_0 = 0$ and $\widetilde{X}_n \geq 0$ $\mathbb{P}$-a.s., so that $\widetilde{X}_n \geq 0$ $\mathbb{Q}$-a.s. (since $\mathbb{P}$ and $\mathbb{Q}$ are equivalent). But since $\mathbb{E}^{\mathbb{Q}}[\widetilde{X}_n] = \widetilde{X}_0 = 0$, it must be the case that $\widetilde{X}_n = 0$, $\mathbb{Q}$-a.s. This implies that $\widetilde{X}_n = 0$ $\mathbb{P}$-a.s., since $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, so there is no arbitrage.

Definition 7.9. A *European contingent claim* with expiration time n is a non-negative $\mathcal{F}_n$-measurable random variable Y, which is called the *payoff* of the claim.

Definition 7.10. A European contingent claim Y is said to be *attainable* (or *hedgeable* or *replicable*) if there exists a constant X_0 and a portfolio process $\pi = (\pi_t)_{t=1}^n$ such that the self-financing wealth process $(X_t)_{t=0}^n$ satisfies

$$X_n(\omega) = Y(\omega), \quad \forall \omega \in \Omega.$$

In this case, for $t = 0, \ldots, n$, we call $V_t := X_t$ the *no-arbitrage price* at time t of Y, and the portfolio which attains $X_n = Y$ is called the *replicating portfolio* for the claim.

Definition 7.11 (Complete market)**.** A financial market is said to be *complete* if every contingent claim is attainable. Otherwise, the market is said to be *incomplete.*

Theorem 7.12 (Second Fundamental Theorem of Asset Pricing (FTAP II))**.** *A finite-state discrete-time arbitrage-free market is complete if and only if there is a unique equivalent martingale measure.*

Here are some examples of European claims in a discrete-time setting with time index set $\mathbb{T} = \{0, 1, \ldots, n\}$.

- A *European call* option, with payoff $Y = (S_n - K)^+$ for fixed strike $K \geq 0$.
- A *European put* option, with payoff $Y = (K - S_n)^+$ for fixed strike $K \geq 0$.
- A *fixed strike lookback call* option, with payoff $Y = (M_n - K)^+$ for fixed strike $K \geq 0$, where M_n is the maximum of the stock price over $\{0, 1, \ldots, n\}$, that is $M_n = \max_{t \in \mathbb{T}} S_t$.
- A *floating strike lookback call* option, with payoff $Y = (S_n - m_n)^+$, where m_n is the minimum of the stock price over $\{0, 1, \ldots, n\}$, that is $m_n = \min_{t \in \mathbb{T}} S_t$.
- An *arithmetic average fixed strike Asian call* option, with payoff $Y = (A_n - K)^+$ for fixed strike $K \geq 0$,, where A_n is the arithmetic average of the stock price over $\{0, 1, \ldots, n\}$, that is
$$A_n = \frac{1}{n+1} \sum_{t=0}^{n} S_t.$$

In a complete market, all contingent claims are attainable. So, given a contingent claim Y, there is a unique trading strategy π with wealth process $X = (X_t)_{t \in \mathbb{T}}$ such that $X_n = Y$ almost surely. This immediately implies that, to avoid arbitrage, the price of the claim at any time $t \in \mathbb{T}$ must be $V_t := X_t$, as in Definition 7.10.

Denote the discount factor from time $t \in \mathbb{T}$ to time zero by D_t. So, with constant interest rate r, $D_t = (1+r)^{-t} = 1/S_t^0$.

Lemma 7.13. *The no-arbitrage price of an attainable claim Y is given by*
$$V_t = \frac{1}{D_t} \mathbb{E}^{\mathbb{Q}}[D_n Y | \mathcal{F}_t], \quad t \in \mathbb{T}. \tag{7.10}$$
Any other price for the claim will lead to an arbitrage opportunity.

Proof. Let $X = (X_t)_{t \in \mathbb{T}}$ be the wealth process of the replicating strategy. The discounted wealth process $\widetilde{X} = DX$ is a $\mathbb{Q}$-martingale, so satisfies
$$\mathbb{E}^{\mathbb{Q}}[D_n X_n | \mathcal{F}_t] = D_t X_t, \quad t \leq n.$$
Using $X_n = Y$ and the definition $V_t := X_t$ yields (7.10).

To show that there is arbitrage if (7.10) is violated, consider buying or selling the claim at time zero (and a similar argument would hold at any time $t \in \mathbb{T}$).

First, suppose $V_0 > \mathbb{E}^{\mathbb{Q}}[D_n Y]$. Sell the claim for V_0 and use the proceeds to invest in the replicating portfolio, which requires an initial investment of $X_0 = \mathbb{E}^{\mathbb{Q}}[D_n Y]$. The wealth in this portfolio at time n is given by $X_n = Y$, by assumption. Therefore, one can, at time zero, invest $V_0 - X_0 > 0$ in the bank, use the proceeds from the replicating portfolio to settle one's obligations from the claim, and make a riskless profit of $(V_0 - X_0)(1+r)^n > 0$. This is an arbitrage.

Similarly, if $V_0 < \mathbb{E}^{\mathbb{Q}}[D_n Y]$, one buys the claim and sells the replicating portfolio, leading to profits $(X_0 - V_0)(1+r)^n > 0$.

□

7.2 Pricing by replication in the binomial model

In this section we shall show that in the binomial model it is possible to replicate any European claim, so the model is complete. If we were to reply on the first FTAP we could deduce this immediately, since we can define a unique EMM $\mathbb{Q}$ via (7.9), repeated below:

$$\mathbb{Q}(\omega_t = \mathrm{H}) = p^{\mathbb{Q}} := \frac{1+r-d}{u-d}, \quad \mathbb{Q}(\omega_t = \mathrm{T}) = q^{\mathbb{Q}} := \frac{u-(1+r)}{u-d}, \tag{7.11}$$

for $t \in \{1, \ldots, n\}$. We will explicity construct a replicating strategy and see that this measure emerges naturally.

7.2.1 Replication in a one-period binomial model

First consider a one-period model, $n = 1$, so $\mathbb{T} = \{0, 1\}$. Suppose an agent sells a claim on the stock at time zero that expires at time 1. There are just two points $\omega \in \Omega$, given by $\omega = \mathrm{H}$ and $\omega = \mathrm{T}$.

The claim pays off an amount Y at time 1, where Y is an $\mathcal{F}_1$-measurable random variable. This measurability condition is relevant; it says that the value of the claim at its maturity date is determined by the coin toss, that is, by the value of the stock price at time 1. This is why it does not make sense to use some stock unrelated to the derivative security in valuing it.

The agent sells the claim at time zero for some price V_0 (to be determined) and attempts to manage the risk from this sale by building a *hedging portfolio* composed of a number $\pi_1 \equiv \pi$ shares of the underlying stock and a number $\pi_1^0 \equiv \pi^0$ of shares of the riskless asset (which has initial value $S_0^0 = 1$).

We suppose that the proceeds from the sale of the claim, V_0, are all that the agent uses to construct a hedging portfolio. Therefore the initial wealth in the hedging portfolio is

$$X_0 = \pi^0 + \pi S_0. \tag{7.12}$$

As the stock price evolves in time the hedging portfolio and option value will also evolve. The option payoff is the random variable $Y(\omega)$ (so for, say, a call option, $Y(\omega) = (S_1(\omega) - K)^+$, where K is the option's strike and $S_1(\omega)$ is the stock price after one coin toss). The agent's hedge portfolio wealth at time 1 is $X_1(\omega)$, given by

$$X_1(\omega) = (1+r)\pi^0 + \pi S_1(\omega).$$

Eliminating π^0 using (7.12), we write X_1 as

$$X_1(\omega) = (1+r)X_0 + \pi(S_1(\omega) - (1+r)S_0).$$

If the hedging portfolio is to successfully manage the risk from the option sale its value must replicate the option payoff in each possible final state, so that we require $X_1(\omega) = Y(\omega)$ for $\omega = \mathrm{H}$ and $\omega = \mathrm{T}$, yielding the equations

$$(1+r)X_0 + \pi(S_1(\mathrm{H}) - (1+r)S_0) = Y(\mathrm{H}), \quad \text{if } \omega = \mathrm{H}, \tag{7.13}$$
$$(1+r)X_0 + \pi(S_1(\mathrm{T}) - (1+r)S_0) = Y(\mathrm{T}), \quad \text{if } \omega = \mathrm{T}. \tag{7.14}$$

Solving these equations for π gives

$$\pi = \frac{Y(\mathrm{H}) - Y(\mathrm{T})}{S_1(\mathrm{H}) - S_1(\mathrm{T})}.$$

Then, the initial wealth is computed from either of (7.13) or (7.14) as

$$X_0 = \frac{1}{1+r}\left[p^{\mathbb{Q}}Y(\mathrm{H}) + q^{\mathbb{Q}}Y(\mathrm{T})\right], \tag{7.15}$$

where we have used (7.11) and $S_1(\mathrm{H}) = uS_0$, $S_1(\mathrm{T}) = dS_0$. (The cash position required can be obtained using (7.12).)

If the agent holds the portfolio (π^0, π), then he will be able to meet all his obligations associated with the claim. Therefore the current claim price is the initial wealth required to do this, or $V_0 = X_0$, as given by (7.15). So we get the claim price at time zero as

$$V_0 = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[Y].$$

The measure $\mathbb{Q}$ is the unique EMM for this one-period market, and is also known as the *risk-neutral probability measure.*

It is clear that $\mathbb{Q}$ is equivalent to the physical measure $\mathbb{P}$, and that $\mathbb{Q}$ is indeed a martingale measure, in that

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{S_1}{1+r}\right] = S_0.$$

It is also clear that the discounted wealth process, and hence the discounted claim price process, is also a $\mathbb{Q}$-martingale, just as we would have expected from the FTAPs.

7.2.2 Generalisation to n-period binomial model

We can easily generalise the above analysis to an n-period model, by simply concatenating a sequence of 1-period models.

Let us place ourselves at some time $t-1$, where $t \in \{1, \ldots, n\}$. Given a fixed outcome $\omega_1 \ldots \omega_{t-1}$ of the first $t-1$ coin tosses, suppose that the values of the stock and a derivative security at time t are $S_t(\omega_t), V_t(\omega_t)$ respectively, if the outcome of the t^{th} coin toss is ω_t (see Figure 7). Then one can trade over $[t-1, t]$ to reproduce the values of the derivative one period later, as follows.

At time $t-1$, after portfolio rebalancing has taken place, the wealth with strategy $(\pi^0, \pi) = (\pi^0_t, \pi_t)_{t=1}^n$ is given by

$$X_{t-1} = \pi^0_t S^0_{t-1} + \pi_t S_{t-1}. \tag{7.16}$$

This evolves to the wealth $X_t(\omega_t)$ at time t, given by

$$X_t(\omega_t) = \pi^0_t(1+r)S^0_{t-1} + \pi_t S_{t-1}\epsilon_t, \quad \epsilon_t = \begin{cases} u, & \omega_t = \mathrm{H}, \\ d, & \omega_t = \mathrm{T}. \end{cases}$$

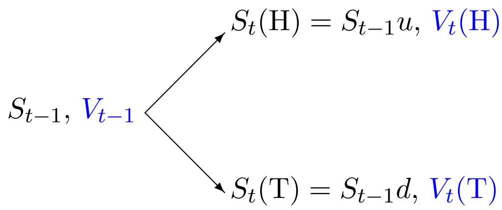


Figure 7: Binomial process for stock price and derivative.

Eliminating $\pi_t^0 S_{t-1}^0$ using (7.16) we get

$$X_t(\omega_t) = (1+r)X_{t-1} + \pi_t S_{t-1}(\epsilon_t - (1+r)), \qquad \omega_t = \mathrm{H}, \mathrm{T}.$$

Writing this out fully as two equations, we have

$$X_t(\mathrm{H}) = (1+r)X_{t-1} + \pi_t S_{t-1}(u - (1+r)), \tag{7.17}$$
$$X_t(\mathrm{T}) = (1+r)X_{t-1} - \pi_t S_{t-1}(1+r-d). \tag{7.18}$$

We require $X_t(\omega_t) = V_t(\omega_t)$, for both $\omega_t = \mathrm{H}$ and $\omega_t = \mathrm{H}$. This requires that the stock holding at the beginning of the interval $[t-1, t)$ must be

$$\pi_t = \frac{V_t(\mathrm{H}) - V)_t(\mathrm{T})}{S_{t-1}(u-d)} = \frac{V_t(\mathrm{H}) - V)_t(\mathrm{T})}{S_t(\mathrm{H}) - S_t(\mathrm{T})}.$$

The required wealth at time $t-1$ is then given from either of (7.17) or (7.18) as

$$X_{t-1} = \frac{1}{1+r}\left[p^{\mathbb{Q}}V_t(\mathrm{H}) + q^{\mathbb{Q}}V_t(\mathrm{T})\right] = \mathbb{E}^{\mathbb{Q}}[(1+r)^{-1}V_t|\mathcal{F}_{t-1}], \quad t = 1, \ldots, n.$$

For no-arbitrage, we must then have that the derivaive value at time $t-1$ must be given by $V_{t-1} = X_{t-1}$:

$$V_{t-1} = \mathbb{E}^{\mathbb{Q}}[(1+r)^{-1}V_t|\mathcal{F}_{t-1}], \quad t = 1, \ldots, n. \tag{7.19}$$

Notice that this implies that the discounted option value $((1+r)^{-t}V_t)_{t=0}^n$ is a $\mathbb{Q}$-martingale (as it should be, since it is replicated by a discounted wealth process which is a $\mathbb{Q}$-martingale.

This shows that one can always find a strategy at any time to reproduce the value of a contingent claim one period later. The key to valuing the contingent claium is thus to begin at the maturitry time and work backwards, computing risk-neutral discoounted expectations. The next section formalises this.

7.3 Completeness of the multiperiod binomial model

The above analysis can clearly be iterated so that in a multiperiod binomial model, we can replicate any contingent claim. The next theorem rigorously demonstrates that a portfolio process to hedge any contingent claim in the binomial model exists, and derives an expression for $\pi_t, t = 1, \ldots, n$.

Define the unique EMM $\mathbb{Q}$ by setting the $\mathbb{Q}$-probability of H on each coin toss is to be $p^{\mathbb{Q}}$, and the $\mathbb{Q}$-probability of T to be $q^{\mathbb{Q}} := 1 - p^{\mathbb{Q}}$, given by (7.11).

Theorem 7.14. *The n-period binomial model is complete. In particular, let Y be European claim with maturity time n, and define*

$$\begin{aligned} V_t(\omega_1 \ldots \omega_t) &:= (1+r)^t \mathbb{E}^{\mathbb{Q}}[(1+r)^{-n} Y | \mathcal{F}_t](\omega_1 \ldots \omega_t), \quad t = 0, \ldots, n, \\ \pi_t(\omega_1 \ldots \omega_{t-1}) &:= \frac{V_t(\omega_1 \ldots \omega_{t-1}\mathrm{H}) - V_t(\omega_1 \ldots \omega_{t-1}\mathrm{T})}{S_t(\omega_1 \ldots \omega_{t-1}\mathrm{H}) - S_t(\omega_1 \ldots \omega_{t-1}\mathrm{T})}, \quad t = 1, \ldots, n. \end{aligned}$$

Then, starting with initial wealth $X_0 := V_0 = \mathbb{E}^{\mathbb{Q}}[(1+r)^{-n} Y]$, the self-financing wealth process corresponding to the portfolio process $\pi_1, \ldots, \pi_n$ is the process $V_0, \ldots, V_n$.

Proof. Let $V_0, \ldots, V_n$ and $\pi_1, \ldots, \pi_n$ be defined as in the theorem. Observe that $V_n = Y$ almost surely.

Start with wealth $X_0 = V_0 = \mathbb{E}^{\mathbb{Q}}[(1+r)^{-n} Y]$ and consider the self-financing value of the process $\pi_1, \ldots, \pi_n$. This wealth satisfies the recursive formula

$$X_{t+1} = (1+r)X_t + \pi_{t+1}(S_{t+1} - (1+r)S_t), \quad t = 0, 1, \ldots, n-1.$$

We need to show that, with X_t, V_t, π_t defined as above, we have

$$X_t = V_t, \quad \text{almost surely}, \quad \forall t \in \{0, \ldots, n\}. \tag{7.20}$$

We proceed by induction. For $t = 0$, (7.20) holds by definition of X_0. Now assume that (7.20) holds for some fixed value of $t \in \{1, \ldots, n-1\}$, i.e. for each fixed $(\omega_1 \ldots \omega_t)$ we have

$$X_t(\omega_1 \ldots \omega_t) = V_t(\omega_1 \ldots \omega_t).$$

Then we need to show that

$$\begin{aligned} X_{t+1}(\omega_1 \ldots \omega_t \mathrm{H}) &= V_{t+1}(\omega_1 \ldots \omega_t \mathrm{H}), \\ X_{t+1}(\omega_1 \ldots \omega_t \mathrm{T}) &= V_{t+1}(\omega_1 \ldots \omega_t \mathrm{T}). \end{aligned}$$

We shall prove the first equality, and note that the second can be proved similarly (an exercise). Note first that $\{(1+r)^{-t} V_t\}_{t=0}^n$ is a martingale under $\mathbb{Q}$, since

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[(1+r)^{-(t+1)} V_{t+1} | \mathcal{F}_t] &= \mathbb{E}^{\mathbb{Q}}[\mathbb{E}^{\mathbb{Q}}[(1+r)^{-n} Y | \mathcal{F}_{t+1}] | \mathcal{F}_t] \quad \text{(defn. of } V_{t+1}) \\ &= \mathbb{E}^{\mathbb{Q}}[(1+r)^{-n} Y | \mathcal{F}_t] \quad \text{(tower property)} \\ &= (1+r)^{-t} V_t. \end{aligned}$$

So in particular,

$$\begin{aligned} V_t(\omega_1 \ldots \omega_t) &= \mathbb{E}^{\mathbb{Q}}[(1+r)^{-1} V_{t+1} | \mathcal{F}_t](\omega_1 \ldots \omega_t) \\ &= \frac{1}{1+r}(p^{\mathbb{Q}} V_{t+1}(\omega_1 \ldots \omega_t \mathrm{H}) + q^{\mathbb{Q}} V_{t+1}(\omega_1 \ldots \omega_t \mathrm{T})). \end{aligned}$$

Since $(\omega_1 \ldots \omega_t)$ will be fixed for the rest of the proof, we simplify notation by suppressing these symbols. For example, the last equation is written as

$$V_t = \frac{1}{1+r}(p^{\mathbb{Q}} V_{t+1}(\mathrm{H}) + q^{\mathbb{Q}} V_{t+1}(\mathrm{T})). \tag{7.21}$$

Now we compute

$$\begin{aligned}
X_{t+1}(\mathrm{H}) &= (1+r)X_t + \pi_{t+1}(S_{t+1}(\mathrm{H}) - (1+r)S_t) \\
&= (1+r)V_t + \pi_{t+1}(S_{t+1}(\mathrm{H}) - (1+r)S_t) \quad (\text{since } X_t = V_t) \\
&= (1+r)V_t + \left[\frac{V_{t+1}(\mathrm{H}) - V_{t+1}(\mathrm{T})}{S_{t+1}(\mathrm{H}) - S_{t+1}(\mathrm{T})}\right](S_{t+1}(\mathrm{H}) - (1+r)S_t) \\
&= p^{\mathbb{Q}}V_{t+1}(\mathrm{H}) + q^{\mathbb{Q}}V_{t+1}(\mathrm{T}) \\
&+ \left[\frac{V_{t+1}(\mathrm{H}) - V_{t+1}(\mathrm{T})}{S_{t+1}(\mathrm{H}) - S_{t+1}(\mathrm{T})}\right](S_{t+1}(\mathrm{H}) - (1+r)S_t) \\
&= p^{\mathbb{Q}}V_{t+1}(\mathrm{H}) + q^{\mathbb{Q}}V_{t+1}(\mathrm{T}) + q^{\mathbb{Q}}(V_{t+1}(\mathrm{H}) - V_{t+1}(\mathrm{T})) \\
&= V_{t+1}(\mathrm{H}),
\end{aligned}$$

where we have used $S_{t+1}(\mathrm{H}) = S_t u$ and $S_{t+1}(\mathrm{T}) = S_t d$.

□

Example 7.15 (European call in 2-period model). Let $u = 2$, $d = 1/u$, $r = 1/4$, $S_0 = 4$, so that $p^{\mathbb{Q}} = q^{\mathbb{Q}} = 1/2$. Consider a European call with expiration time 2 and payoff function $Y = (S_2 - 5)^+$. The possible stock prices in this model are shown in Figure 8.

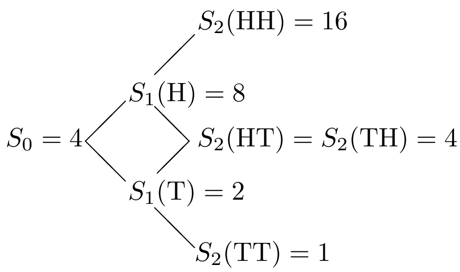


Figure 8: Two period binomial lattice

There are four elements $\omega \in \Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$, so in principle there are four possible final stock prices. But in fact, two of the outcomes ω lead to the same stock price. We say that the stock price is *path-independent* since it only depends on the *number* of H and T in the sequence $\omega = (\omega_1, \omega_2)$ (where $\omega_t, t = 1, 2$ is either H or T), and does not depend on the order in which the H and T occur. Thus $S_2(\mathrm{HT}) = S_2(\mathrm{TH}) = 4$. for example. The terminal option payoffs for each $\omega \in \Omega$ are

$$Y(\mathrm{HH}) = 11, \quad Y(\mathrm{HT}) = Y(\mathrm{HT}) = Y(\mathrm{TH}) = Y(\mathrm{TT}) = 0,$$

and these are of course the option values at time 2:

$$V_2(\mathrm{HH}) = 11, \quad V_2(\mathrm{HT}) = V_2(\mathrm{HT}) = V_2(\mathrm{TH}) = V_2(\mathrm{TT}) = 0.$$

Then using the binomial algorithm in Theorem 7.14 we "work backwards in time" using the

fact that V is a $\mathbb{Q}$-martingale, to obtain

$$\begin{aligned}
V_1(\mathrm{H}) &= \frac{1}{1+r}(p^{\mathbb{Q}}V_2(\mathrm{HH}) + q^{\mathbb{Q}}V_2(\mathrm{HT})) = \frac{4}{5}\left(\frac{1}{2}(11) + \frac{1}{2}(0)\right) = \frac{22}{5}, \\
V_1(\mathrm{T}) &= \frac{1}{1+r}(p^{\mathbb{Q}}V_2(\mathrm{TH}) + q^{\mathbb{Q}}V_2(\mathrm{TT})) = \frac{4}{5}\left(\frac{1}{2}(0) + \frac{1}{2}(0)\right) = 0, \\
V_0 &= \frac{1}{1+r}(p^{\mathbb{Q}}V_1(\mathrm{H}) + q^{\mathbb{Q}}V_1(\mathrm{T})) = \frac{4}{5}\left(\frac{1}{2}\left(\frac{22}{5}\right) + \frac{1}{2}(0)\right) = \frac{44}{25} = 1.76.
\end{aligned}$$

8 American options in the binomial model

We briefly discuss the pricing of *American* derivative securities in the binomial model. American derivative securities can be exercised at any time prior to maturity.

Definition 8.1. In a discrete-time framework with time set $\mathbb{T} = \{0, 1, \dots, n\}$, an *American derivative security* with maturity n is a sequence of nonnegative random variables $(Y_t)_{t=0}^n$ such that for each $t \in \mathbb{T}$, Y_t is $\mathcal{F}_t$-measurable. The owner of an American derivative security can exercise at any time $t \in \mathbb{T}$, and if he does, he receives the payment Y_t.

For example, an American put option of strike K on a stock price $S = (S_t)_{t=0}^n$ can be exercised at any time $t \in \mathbb{T}$ to give the owner a payment $Y_t := (K - S_t)^+$, which is called the *intrinsic value* of the option at time t.

Recall the pricing of European securities. Consider a binomial model with n periods, so the time set is $\mathbb{T} = \{0, 1, \dots, n\}$. Suppose Y_n is the payoff of a European derivative. For $t \in \mathbb{T}$, we define by backward recursion

$$V_n := Y_n, \quad V_t := \frac{1}{1+r}[p^{\mathbb{Q}}V_{t+1}(\mathrm{H}) + q^{\mathbb{Q}}V_{t+1}(\mathrm{T})], \quad t = 0, \dots, n-1, \tag{8.1}$$

where, as before, the second equation is a shorthand for

$$\begin{aligned}
V_t(\omega_1 \dots \omega_t) &= \mathbb{E}^{\mathbb{Q}}[(1+r)^{-1}V_{t+1}|\mathcal{F}_t](\omega_1 \dots \omega_t) \\
&= \frac{1}{1+r}(p^{\mathbb{Q}}V_{t+1}(\omega_1 \dots \omega_t\mathrm{H}) + q^{\mathbb{Q}}V_{t+1}(\omega_1 \dots \omega_t\mathrm{T})).
\end{aligned}$$

Then V_t is the value of the option at time $t \in \mathbb{T}$, and the hedging portfolio over $[t-1, t)$ is π_t given by

$$\pi_t = \frac{V_t(\mathrm{H}) - V_t(\mathrm{T})}{S_t(\mathrm{H}) - S_t(\mathrm{T})} = \frac{V_t(\mathrm{H}) - V_t(\mathrm{T})}{S_{t-1}(u-d)}, \quad t = 1, \dots, n,$$

which is shorthand for

$$\pi_t(\omega_1 \dots \omega_{t-1}) = \frac{V_t(\omega_1 \dots \omega_{t-1}\mathrm{H}) - V_t(\omega_1 \dots \omega_{t-1}\mathrm{T})}{S_t(\omega_1 \dots \omega_{t-1}\mathrm{H}) - S_t(\omega_1 \dots \omega_{t-1}\mathrm{T})}, \quad t = 1, \dots, n.$$

Now suppose the option is American, with payoff $Y = (Y_t)_{t=0}^n$. At any time $t \in \mathbb{T}$, the holder of the American derivative can exercise the option and receive the payment Y_t. Hence, the hedging portfolio should create a wealth process X which satisfies

$$X_t \geq Y_t, \quad \forall \quad t \in \mathbb{T}, \quad \text{almost surely.}$$

This is because the value of the American option at time t is at least as much as the so-called *intrinsic value* Y_t, and the value of the hedging portfolio at that time must equal the value of the option.

This suggests that, to price an American derivative, we should replace the European algorithm (8.1) by to the following *American algorithm*:

$$V_n = Y_n, \quad V_t = \max\left[Y_t, \frac{1}{1+r}[p^{\mathbb{Q}}V_{t+1}(\mathrm{H}) + q^{\mathbb{Q}}V_{t+1}(\mathrm{T})]\right], \quad t = 0, \dots, n-1, \tag{8.2}$$

which checks whether the intrinsic value is greater than the value of the discounted risk-neutral expectation, which would signify that the option would be exercised in that state. Then V_t would be the value of the American derivative at time $t \in \mathbb{T}$.

Remark 8.2 (Super-martingale property of American option price)*.* In valuing European options we found that the discounted option value is a $\mathbb{Q}$-martingale (recall, for example, (7.19)). From (8.2). We see that the value of the American option can be *greater* than that given by a discounted risk-neutral expectation, because of the possibility of early exercise. In other words, we might have that

$$V_t > \mathbb{E}^{\mathbb{Q}}\left[\frac{1}{1+r}V_{t+1}\middle|\mathcal{F}_t\right],$$

or, equivalently

$$\mathbb{E}^{\mathbb{Q}}[(1+r)^{-(t+1)}V_{t+1}|\mathcal{F}_t] < (1+r)^{-t}V_t,$$

so that the option value is a $\mathbb{Q}$-*supermartingale.* (It turns out that the value process of an American option is the smallest supermartingale that dominates the payoff, though we do not prove this here.)

Example 8.3 (American put in a 2-period model)*.* Consider an American put option in a 2-period binomial model with $u = 2$, $d = 1/u$, $r = 1/4$, $S_0 = 4$, so that $p^{\mathbb{Q}} = q^{\mathbb{Q}} = 1/2$. Let the option have payoff function $Y_t = (5 - S_t)^+$. The possible stock prices in this model are shown in Figure 9. The terminal values of the option are given by $V_2 = Y_2 = (5 - S_2)^+$ and these are also shown in the figure.

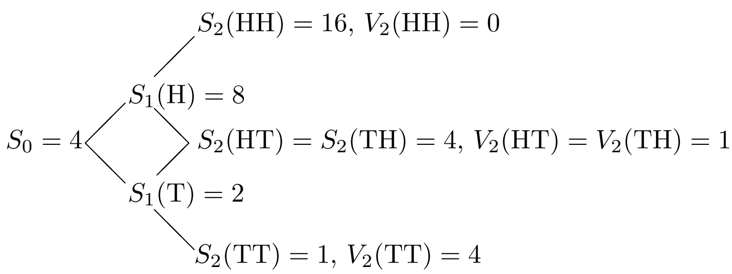


Figure 9: Stock price and terminal value of American put

Then the values of the option at time 1 are:

$$\begin{aligned} V_1(\mathrm{H}) &= \max\left[(5-8)^+, \frac{4}{5}\left(\frac{1}{2}\cdot 0 + \frac{1}{2}\cdot 1\right)\right] = \max\left[0, \frac{2}{5}\right] = \frac{2}{5} \\ V_1(\mathrm{T}) &= \max\left[(5-2)^+, \frac{4}{5}\left(\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 4\right)\right] = \max[3,2] = 3. \end{aligned}$$

In particular, we notice that at time 1, and for $\omega_1 = \mathrm{T}$, the option should be exercised, as the intinsic value is greater than the discounted risk-neutral expectation of later values.

The option value at time zero is

$$V_0 = \max\left[(5-4)^+, \frac{4}{5}\left(\frac{1}{2}\left(\frac{2}{5}\right) + \frac{1}{2}(3)\right)\right] = \max\left[1, \frac{34}{25}\right] = \frac{34}{25} = 1.36.$$

Now let us attempt to construct the hedging portfolio for this option. We begin with initial wealth $X_0 = 34/25$, and we compute π_1 via the replication condition for $\omega_1 = \mathrm{H}$:

$$X_1(\mathrm{H}) = (1+r)X_0 + \pi_1(S_1(\mathrm{H}) - (1+r)S_0) = V_1(\mathrm{H}) = \frac{2}{5},$$

which yields $\pi_1 = -13/30$. We could just as well calculate π_1 by looking at the wealth $X_1(\mathrm{T})$, as follows:

$$X_1(\mathrm{T}) = (1+r)X_0 + \pi_1(S_1(\mathrm{T}) - (1+r)S_0) = V_1(\mathrm{T}) = 3,$$

which also yields $\pi_1 = -13/30$. Now let us try to compute π_2 in a similar manner:

$$X_2(\mathrm{HH}) = (1+r)X_1(\mathrm{H}) + \pi_2(\mathrm{H})(S_2(\mathrm{HH}) - (1+r)S_1(\mathrm{H})) = V_2(\mathrm{HH}) = 0,$$

which yields $\pi_2(\mathrm{H}) = -\frac{1}{12}$. The same result is obtained if one considers the wealth $X_2(\mathrm{HT})$. Now let us try to compute $\pi_2(\mathrm{T})$ as follows:

$$X_2(\mathrm{TH}) = (1+r)X_1(\mathrm{T}) + \pi_2(\mathrm{T})(S_2(\mathrm{TH}) - (1+r)S_1(\mathrm{T})) = V_2(\mathrm{TH}) = 1,$$

which yields $\pi_2(\mathrm{T}) = -11/6$. However, if we try to compute $\pi_2(\mathrm{T})$ using $X_2(\mathrm{TT})$, we get

$$X_2(\mathrm{TT}) = (1+r)X_1(\mathrm{T}) + \pi_2(\mathrm{T})(S_2(\mathrm{TT}) - (1+r)S_1(\mathrm{T})) = V_2(\mathrm{TT}) = 4,$$

which yields $\pi_2(\mathrm{T}) = -1/6$. In other words, we get different answers for $\pi_2(\mathrm{T})$, the position in stock that should be chosen at the start of the interval $[1,2)$ when $\omega_1 = \mathrm{T}$! This apparent anomaly has arisen because $X_1(\mathrm{T}) = 3$ (since the American put is exercised when $\omega_1 = \mathrm{T}$) rather than 2, which would be the case if the option were European (and you can check that in this case the above calculations would both have yielded $\pi_2(\mathrm{T}) = -1$).

This example showns that we need to analyse the hedging portfolio for an American option more closely.

8.1 Value of hedging portfolio for an American option

Consider the following generalisation of the evolution of the wealth of a self-financing portfolio, equation (7.6):

$$X_t = (1+r)(X_{t-1} - C_{t-1}) + \pi_t(S_t - (1+r)S_{t-1}), \quad t = 1, \ldots, n, \tag{8.3}$$

where, for $t \in \{0, 1, \ldots, n-1\}$, C_t represents the amount of wealth *consumed* at time t. In other words, we are allowing for some funds to be withdrawn from the self-financing portfolio. We found earlier that, for a self-financing portfolio, the discounted wealth process $((1+r)^{-t}X_t)_{t=0}^T$ is a martingale. The consequence of allowing consumption from the portfolio will mean that the discounted wealth process will be a *supermartingale* (i.e. it will tend to go *down*).

To appreciate why this adjustment might be needed, consider the American algorithm in (8.2). We see that the value of the option can be *greater* than that given by a discounted risk-neutral expectation, because of the possibility of early exercise. In other words, we might have that

$$V_t > \mathbb{E}^{\mathbb{Q}}\left[\left.\frac{1}{1+r}V_{t+1}\right|\mathcal{F}_t\right],$$

or, equivalently

$$\mathbb{E}^{\mathbb{Q}}[(1+r)^{-(t+1)}V_{t+1}|\mathcal{F}_t] < (1+r)^{-t}V_t,$$

so that the option value is a supermartingale. (It turns out that the value process of an American option is the smallest supermartingale that dominates the payoff, though we do not prove this here.)

To see how consumption enters the hedging portfolio, consider the situation in which

$$V_t > \mathbb{E}^{\mathbb{Q}}\left[\left.\frac{1}{1+r}V_{t+1}\right|\mathcal{F}_t\right]. \tag{8.4}$$

Then the holder of the American option should exercise (this is the case in the state $\omega_1 = \mathrm{T}$ in Example 8.3), so that hedging should stop at this point (which is why we had difficulty isolating what the hedging portfolio should be in the example). If the holder of the option does *not* exercise, then the seller of the option may consume to close the gap between the left and right hand sides of (8.4). By doing this, he can ensure that $X_t = V_t$ for all $t \in \mathbb{T}$, where V_t is the value defined by the American algorithm.

In Example 8.3, we had $V_1(\mathrm{T}) = 3$, $V_2(\mathrm{TH}) = 1$, $V_2(\mathrm{TT}) = 4$, so that

$$\mathbb{E}^{\mathbb{Q}}\left[\left.\frac{1}{1+r}V_2\right|\mathcal{F}_1\right](\mathrm{T}) = \frac{4}{5}\left[\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 4\right] = 2,$$

and there is a gap if size 1 in (8.4). If the owner of the option does not exercise it at time 1 in the state $\omega_1 = \mathrm{T}$, then the seller can consume an amount 1 at time 1. Thereafter he uses the usual hedging portfolio

$$\pi_t = \frac{V_t(\mathrm{H}) - V_t(\mathrm{T})}{(u-d)S_{t-1}}.$$

In the example, we had $V_1(\mathrm{T}) = Y_1(\mathrm{T})$, which means that, acting optimally, the holder of the option should exercise. It turns out that it is optimal for the owner of the American option to exercise whenever its value V_t agrees with its intrinsic value Y_t.

Part III
Continuous time

9 Brownian motion

9.1 Random walk

Toss a coin infinitely many times, so that the sample space Ω is the set of all infinite sequences $\omega = (\omega_1\omega_2\ldots)$ of H and T. One can construct a well-defined probability space $(\Omega, \mathcal{F}, \mathbb{P})$ called the space of infinite coin tosses (though this is not completely trivial, as Ω is an uncountably infinite space), as well as a filtration $(\mathcal{F})_{t\geq 0}$ on this space. We do not have time to delve into the construction of infinite coin toss space here. Chapters 1 and 2 of Shreve [13] has a detailed account.

Assume that each toss is independent, and that on each toss the probability of H is p, so that the probability of T is $q := 1 - p$. Define

$$Y_j(\omega) := \begin{cases} \alpha & \text{if } \omega_j = \mathrm{H}, \\ \beta & \text{if } \omega_j = \mathrm{T}. \end{cases} \qquad j = 1, 2, \ldots \tag{9.1}$$

The random variable Y_j, which always takes one of two values, is sometimes called a *Bernoulli* random variable.

Define a process $M = (M_k)_{k=0}^{\infty}$ by

$$\begin{aligned} M_0 &:= 0, \\ M_k &:= \sum_{j=1}^{k} Y_j, \quad k = 1, 2, \ldots. \end{aligned} \tag{9.2}$$

The process $(M_k)_{k=0}^{\infty}$ is called a *random walk*. It is the sum of independent, identically distributed (i.i.d.) Bernoulli variables, and is sometimes called a *binomial* random variable.

Remark 9.1 (Symmetric random walk). For $\alpha = 1, \beta = -1, p = q = \frac{1}{2}$, the process $(M_k)_{k=0}^{\infty}$ is a *symmetric random walk*, whose analogue in continuous time is *Brownian motion*, as we shall see.

9.2 BM as scaled limit of symmetric random walk

On the infinite coin toss space $(\Omega, \mathcal{F}, \mathbb{P})$, define the random variables

$$X_j(\omega) := \left\{ \begin{array}{ll} 1 & \text{if } \omega_j = \mathrm{H}, \\ -1 & \text{if } \omega_j = \mathrm{T}, \end{array} \right\}, \quad j = 1, 2 \ldots$$

with $\mathbb{P}\{\omega_j = \mathrm{H}\} = \mathbb{P}\{\omega_j = \mathrm{T}\} = \frac{1}{2}$, so that each X_j has mean zero and variance 1, (i.e. $\mathbb{E}[X_j|\mathcal{F}_{j-1}] = 0$ and $\mathbb{E}[X_j^2|\mathcal{F}_{j-1}] = 1$). So $X_1, X_2, \ldots$ is a sequence of independent, identically distributed random variables. Then define the symmetric random walk M via

$$\begin{aligned} M_0 &:= 0, \\ M_k &:= \sum_{j=1}^{k} X_j, \quad k = 1, 2, \ldots. \end{aligned}$$

By the Law of Large Numbers we know that

$$\frac{1}{k} M_k \to 0, \quad \text{almost surely, as } k \to \infty.$$

By the Central Limit Theorem we know that for large k, $M_k/\sqrt{k}$ is approximately standard normal:

$$\frac{1}{\sqrt{k}} M_k \to Z \sim N(0,1), \quad \text{almost surely, as } k \to \infty.$$

Brownian motion arises if we suitably speed up the tossing of the coins and scale the size of each random walk increment. To this end, if $t \geq 0$ is of the form $t = k/n =: k\delta t$ (so $\delta t = 1/n$ is the time betwen coin tosses) for positive integers k, n, then define a continuous time process via

$$W_t^{(n)} := \frac{1}{\sqrt{n}} M_{nt} = \sqrt{\frac{t}{k}} M_k = \sqrt{\delta t} M_{t/\delta t}, \quad t \geq 0,$$

with linear interpolation used to define $W_t^{(n)}$ for any times $t \geq 0$ not of the form k/n. Take the limit $k \to \infty$, with t fixed.[3] Then since $\frac{1}{\sqrt{k}} M_k \to Z \sim N(0,1)$ as $k \to \infty$, we have that

$$W_t^{(n)} \to W_t \sim N(0,t), \quad \text{as } n \to \infty,$$

and we call the process $(W_t)_{t \geq 0}$ a *standard Brownian motion.*

Notice that with $t = k\delta t$, we have (though this is purely formal)

$$\frac{\mathrm{d}W_t^{(n)}}{\mathrm{d}t} = \lim_{\delta t \to 0} \frac{W_{t+\delta t}^{(n)} - W_t^{(n)}}{\delta t} = \lim_{\delta t \to 0} \frac{1}{\sqrt{\delta t}} X_{k+1} \to \infty.$$

If, instead of $W_t^{(n)}$, we were to define

$$V_t^{(n)} := \frac{1}{n} M_{nt} = \frac{t}{k} M_k = \delta t M_{t/\delta t},$$

then by the Law of Large Numbers,

$$V_t^{(n)} \to 0, \quad \text{as } k \to \infty,$$

[3] Equivalently, $n \to \infty$, or equivalently, $\delta t \to 0$, where $\delta t := 1/n$ is the time interval between coin tosses. Then $k = nt = t/\delta t$, and hence $t = k\delta t$, so we are speeding up the coin tossing, and since $t/k = \delta t$, $W_t^{(n)} = \sqrt{\delta t} M_{t/\delta t} = \sqrt{\delta t} M_k$, so that we are scaling each increment of the random walk by $\sqrt{\delta t}$.

and

$$\frac{\mathrm{d}V_t^{(n)}}{\mathrm{d}t}(t_j) = \lim_{\delta t \to 0} \frac{V_{t+\delta t}^{(n)} - V_t^{(n)}}{\delta t} = \lim_{\delta t \to 0} X_{k+1} = \pm 1,$$

so while the derivative of $V^{(n)}$ is defined (unlike that of $W^{(n)}$), the process $V^{(n)}$ is trivially zero in the limit.

In other words the Brownian "particle" can only have motion if it has infinite velocity. This is a manifestation of the fact that paths of W_t are almost surely continuous but not differentiable, as we will see again in a short while.

Remark 9.2 (Random walks and the binomial model). In the binomial model, the logarithm of the stock price process follows a random walk. A similar analysis as above can be used to show that the continuous time limit of a binomial model has stock price process given by

$$S_t = S_0 \exp\left[\left(b - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right], \quad t \geq 0, \tag{9.3}$$

where where b and $\sigma > 0$ are constants related to the binomial parameters u, d and to the probability p of the stock price rising in the binomial tree, by

$$\log u = \left(b - \frac{1}{2}\sigma^2\right)\delta t + \sigma\sqrt{\delta t}, \quad \beta = \left(b - \frac{1}{2}\sigma^2\right)\delta t - \sigma\sqrt{\delta t}, \quad p = \frac{1}{2},$$

The process (9.3) is known as *geometric Brownian motion*.

9.3 Brownian motion

We shall see that Brownian motion (BM) is a continuous stochastic process which is Markov, Gaussian, and a martingale.

Let $X \sim N(\mu, \sigma^2)$ denote that a random variable X is normally distributed with mean μ and variance σ^2.

Definition 9.3 (Brownian motion). A standard 1-dimensional *Brownian motion* (BM) is a continuous adapted process $W := (W_t, \mathcal{F}_t)_{0 \leq t < \infty}$ on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ with the properties that $W_0 = 0$ a.s. and, for $0 \leq s < t$, $W_t - W_s$ is independent of $\mathcal{F}_s$ and normally distributed as $W_t - W_s \sim N(0, t - s)$.

The filtration $(\mathcal{F}_t)_{t \geq 0}$ is a part of the definition of BM. However, if we are given $(W_t)_{t \geq 0}$ but no filtration, and if we know that W has stationary, independent increments and that $W_t = W_t - W_0 \sim N(0, t)$, then with $(\mathcal{F}_t^W)_{t \geq 0}$ being the filtration generated by the BM, we have $(W_t, \mathcal{F}_t^W)_{0 \leq t < \infty}$ is a BM in the sense of Definition 9.3 (see ([9], Problem 1.4).

Here is another definition BM, based on a process called *quadratic variation* , one definition of which is given below. Let $\mathcal{M}_2$ denote the space of right-continuous square-integrable martingales on a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$: that is, for $M := (M_t)_{t \geq 0} \in \mathcal{M}_2$ we have $M_0 = 0$ a.s., and $\mathbb{E}[M_t^2] < \infty$, for all $t \geq 0$.

Definition 9.4 (Quadratic variation). For $M \in \mathcal{M}_2$, the *quadratic variation* (QV) of M is the unique, increasing adapted process $[M]$ such that $[M]_0 = 0$ a.s. and such that $(M_t^2 - [M]_t)_{t \geq 0}$ is a martingale.

Definition 9.5 (Cross-variation)**.** For $X, Y \in \mathcal{M}_2$, define their cross-variation process $([X,Y]_t)_{t\geq 0}$ by

$$[X,Y]_t := \frac{1}{4}\left([X+Y]_t - [X-Y]_t\right),$$

For $X, Y \in \mathcal{M}_2^c$ (i.e. continuous), this is the unique increasing adapted process $[X,Y]$ such that $[X,Y]_0 = 0$ a.s. and such that $((XY - [X,Y])_t)_{t\geq 0}$ is a martingale.

Remark 9.6. For Brownian motion $W := (W_t)_{t\geq 0}$, we have $[W]_t = t$, since $W_t^2 - t$ is a martingale (see Problem Sheet 2). Indeed, Brownian motion may be defined as the unique continuous process that satisfies this property.

We denote by $(\mathcal{F}_t)_{t\geq 0}$ the filtration generated by Brownian motion. Its required properties are:

- For each t, W_t is $\mathcal{F}_t$-measurable;
- for each t and for $t < t_1 < t_2 < \ldots < t_n$, the Brownian motion increments $W_{t_1} - W_t, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent of $\mathcal{F}_t$.

Here is one way to construct $\mathcal{F}_t$. First fix t. Let $s \in [0,t]$ and $A \in \mathcal{B}(\mathbb{R})$ be given. Put the set

$$\{W_s \in A\} = \{\omega : W_s(\omega) \in A\}$$

in $\mathcal{F}_t$. Do this for all possible numbers $s \in [0,t]$ and all Borel sets $A \in \mathcal{B}(\mathbb{R})$. Then put in every other set required by the σ-algebra properties. This σ-algebra $\mathcal{F}_t$ contains exactly the information learned by observing the Brownian motion up to time t, and $(\mathcal{F}_t)_{t\geq 0}$ is called the filtration generated by the Brownian motion.

9.4 Properties of BM

Stationarity We say a stochastic process $X = (X_t)_{t\geq 0}$ is stationary if X_t has the same distribution as X_{t+h} for any $h > 0$. Brownian motion has *stationary increments.* To see this, define the increment process $I = (I_t)_{t\geq 0}$ by $I_t := W_{t+h} - W_t$. Then $I_t \sim N(0,h)$, and $I_{t+h} = W_{t+2h} - W_{t+h} \sim N(0,h)$ have the same distribution. This is equivalent to saying that the process $(W_{t+h} - W_t)_{h\geq 0}$ has the same distribution for all t.

Martingale property The independent increments property allows us to shw that BM is a martingale. For $0 \leq s \leq t$ we have

$$\mathbb{E}[W_t|\mathcal{F}_s] = \mathbb{E}[W_t - W_s + W_s|\mathcal{F}_s] = \mathbb{E}[W_t - W_s|\mathcal{F}_s] + W_s = \mathbb{E}[W_t - W_s] + W_s = W_s.$$

Covariance of BM at different times Let $0 \leq s \leq t$ be given. Then W_s and $W_t - W_s$ are independent, and (W_s, W_t) are jointly normal with $\mathbb{E}[W_s] = \mathbb{E}[W_t] = \mathbb{E}[W_t - W_s] = 0$,

$\text{var}(W_s) = s$, $\text{var}(W_t) = t$, $\text{var}(W_t - W_s) = t - s$, so that the covariance of W_s and W_t is

$$\begin{aligned}
\text{cov}(W_s, W_t) &:= \mathbb{E}[(W_s - \mathbb{E}[W_s])(W_t - \mathbb{E}[W_t])] \\
&= \mathbb{E}[W_s W_t] \\
&= \mathbb{E}[W_s(W_t - W_s + W_s)] \\
&= \mathbb{E}[W_s(W_t - W_s)] + \mathbb{E}[W_s^2] \\
&= \mathbb{E}[W_s]\mathbb{E}[W_t - W_s] + s \quad \text{(by independence)} \\
&= s.
\end{aligned}$$

Thus, for any $s \geq 0$, $t \geq 0$ (not necessarily $s \leq t$), we have

$$\text{cov}(W_s, W_t) = \mathbb{E}[W_s W_t] = s \wedge t = \min(s, t),$$

or, equivalently, the covariance matrix of the vector $\mathbf{W}_{s,t} = (W_s, W_t)$ is $C \equiv A^{-1}$, given by

$$C = A^{-1} = \begin{pmatrix} s & s \wedge t \\ s \wedge t & t \end{pmatrix}, \quad \text{(positive definite, symmetric).}$$

Definition 9.7 (Transition density)**.** Fix $x \in \mathbb{R}$, $t_0 \in \mathbb{R}_+$. Then

$$\mathbb{P}(W_{t_0+t} \in [y, y + \,\mathrm{d}y] | W_{t_0} = x) = p(t, x, y)\,\mathrm{d}y,$$

where the *transition density* of Brownian motion is the function

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right), \quad y \in \mathbb{R}, \quad t > 0.$$

This is the probability density that the BM moves from x to $y \in \mathbb{R}$ in a time period t.

Starting points other than zero For a standard Brownian motion W that starts at zero we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that satisfies $\mathbb{P}\{W_0 = 0\} = 1$. Then for $t \geq 0$, $W_t \sim N(0, t)$. For $x \in \mathbb{R}$, we can define a process $W_t^x := x + W_t$ which will satisfy $\mathbb{P}\{W_0^x = x\} = 1$ and, for $t \geq 0$, $W_t^x \sim N(x, t)$.

Equivalently, we can define another probability measure $\mathbb{P}^x$ (or, more formally, a probability space $(\Omega, \mathcal{F}, \mathbb{P}^x)$) under which $\mathbb{P}^x\{W_0 = x\} = 1$, and with W having stationary independent increments under $\mathbb{P}^x$: for $s \leq t$, $W_t - W_s \sim N(0, t - s)$ and independent of $\mathcal{F}_s$. Then, under $\mathbb{P}^x$, $W_t \sim N(x, t)$. In this case, we say that W is a *Brownian motion starting at* x. We see that such a Brownian motion is equivalent to $x + W$, where W is a standard Brownian motion starting at zero.

Note that:

- If $x \neq 0$, then $\mathbb{P}^x$ puts all its probability on a completely different set from $\mathbb{P}$.
- The distribution of W_t under $\mathbb{P}^x$ is the same as the distribution of $W_t^x = x + W_t$ under $\mathbb{P}$, that is
$$\text{Law}(W^x, \mathbb{P}) = \text{Law}(W, \mathbb{P}^x).$$

Markov property We can show that W is a Markov process as follows. Recall that the Markov property is equivalent to stating that for $s \geq 0$, $t \geq 0$, we have $\mathbb{E}[h(W_{s+t})|\mathcal{F}_s] = g(W_s)$, where h and g are functions. Consider

$$\mathbb{E}[h(W_{s+t})|\mathcal{F}_s] = \mathbb{E}[h(W_{s+t} - W_s + W_s)|\mathcal{F}_s].$$

Use the properties that $W_{s+t} - W_s$ is independent of $\mathcal{F}_s$, and that W_s is $\mathcal{F}_s$-measurable, along with the following *independence lemma*: if X, Y are random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and if $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, with

- X $\mathcal{G}$-measurable
- Y independent of $\mathcal{G}$.

Then if $f(x, y)$ is a function of two variables, and if we define

$$g(x) := \mathbb{E}[f(x, Y)],$$

then we have

$$\mathbb{E}[f(X, Y)|\mathcal{G}] = g(X).$$

In this lemma, take $\mathcal{G} = \mathcal{F}_s$, $X = W_s$, $Y = W_{s+t} - W_s$, and $f(x, y) = h(x + y)$. Then define

$$\begin{aligned} g(x) &:= \mathbb{E}[h(W_{s+t} - W_s + x)] \\ &= \mathbb{E}[h(x + W_t) \quad \text{(since } W_t \sim N(0, t) \text{ has the same distribution as } W_{s+t} - W_s) \\ &= \mathbb{E}^x[h(W_t)]. \end{aligned}$$

Then

$$\mathbb{E}[h(W_{s+t})|\mathcal{F}_s] = g(W_s) = \mathbb{E}^{W_s}[h(W_t)],$$

which is the Markov property.

In fact, Brownian motion has the *strong Markov property* (though we do not prove this).

Strong Markov property Fix $x \in \mathbb{R}$ and define

$$\tau := \min\{t \geq 0 | W_t = x\}.$$

Then we have

$$\mathbb{E}[h(W_{\tau+t})|\mathcal{F}_\tau] = g(x) = \mathbb{E}^x h(W_t).$$

9.5 Quadratic variation of BM

Definition 9.8 (p^{th} variation)**.** Let $\mathcal{P} = \{t_0, t_1, \ldots, t_n\}$ be a *partition* of $[0, t]$, i.e.

$$0 = t_0 \leq t_1 \leq \ldots \leq t_n = t.$$

The *mesh* of the partition is defined to be

$$\|\mathcal{P}\| = \max_{k=0,\ldots,n-1} |t_{k+1} - t_k|$$

The p^{th} *variation* of a function $f : \mathbb{R}^+ \to \mathbb{R}$ on an interval $[0, t]$, $[f, f]^{(p)}(t)$, is defined by

$$[f, f]_t^{(p)} := \lim_{\|\mathcal{P}\| \to 0} \sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)|^p. \tag{9.4}$$

In particular, if $p = 1$ this is called the *total variation* (or the *first variation*) and if $p = 2$ this is called the *quadratic variation*.

9.5.1 First variation

Consider the *first variation* (or *total variation*), $[f, f]_t^{(1)}$, of a function f. Suppose f is differentiable. Then the Mean Value Theorem[4] implies that in each subinterval $[t_k, t_{k+1}]$, there is a point t_k^* such that

$$f(t_{k+1}) - f(t_k) = (t_{k+1} - t_k) f'(t_k^*).$$

Then

$$\sum_{k=0}^{n-1} |f(t_{k+1} - f(t_k)| = \sum_{k=0}^{n-1} |f'(t_k^*)|(t_{k+1} - t_k),$$

and so

$$\begin{aligned} [f, f]_t^{(1)} &= \lim_{\|\mathcal{P}\| \to 0} \sum_{k=0}^{n-1} |f'(t_k^*)|(t_{k+1} - t_k) \\ &= \int_0^t |f'(s)| \, \mathrm{d}s. \end{aligned}$$

Thus, first variation measures the total amount of up and down motion of the path of f over the interval $[0, t]$.

9.5.2 Quadratic variation of Brownian motion

To simplify notation, we write $[f, f]_t^{(2)} = [f]_t$ for the quadratic variation of a function f over the interval $[0, t]$.

Lemma 9.9. *If f is differentiable, then $[f]_t = 0$.*

Proof.

$$\begin{aligned} \sum_{k=0}^{n-1} |f(t_{k+1} - f(t_k)|^2 &= \sum_{k=0}^{n-1} |f'(t_k^*)|^2 (t_{k+1} - t_k)^2 \\ &\leq \|\mathcal{P}\| \sum_{k=0}^{n-1} |f'(t_k^*)|^2 (t_{k+1} - t_k), \end{aligned}$$

[4]The Mean Value Theorem states that if f is differentiable in (a, b), then there is a point $x \in (a, b)$ at which $f(b) - f(a) = (b - a) f'(x)$.

and so

$$\begin{aligned}
[f]_t &\leq \lim_{\|\mathcal{P}\|\to 0} \|\mathcal{P}\| \lim_{\|\mathcal{P}\|\to 0} \sum_{k=0}^{n-1} |f'(t_k^*)|^2 (t_{k+1} - t_k) \\
&= \lim_{\|\mathcal{P}\|\to 0} \|\mathcal{P}\| \int_0^t |f'(s)|^2 \, \mathrm{d}s \\
&= 0.
\end{aligned}$$

□

Theorem 9.10. *For Brownian motion $W = (W_t)_{t\geq 0}$ we have*

$$[W]_t = t, \quad t \geq 0,$$

or more precisely

$$\mathbb{P}\{\omega \in \Omega : [W]_t(\omega) = t\} = 1.$$

In particular, the paths of Brownian motion are not differentiable.

Some words of intuition. Since $W_t \sim N(0,t)$ its moment generating function $M(a)$ is given by

$$M(a) = \mathbb{E}[\exp(aW_t)] = \exp\left(\frac{1}{2}a^2 t\right).$$

Expanding the exponentials as Taylor series in powers of a yields $\mathbb{E}[W_t] = 0$, $\mathbb{E}[W_t^2] = t$, $\mathbb{E}[W_t^3] = 0$, $\mathbb{E}[W_t^4] = 3t^2$. Hence the variance of W_t^2 is

$$\mathrm{var}[W_t^2] = \mathbb{E}[W_t^4] - (\mathbb{E}[W_t^2])^2 = 3t^2 - t^2 = 2t^2.$$

The important observation is that, for small t, the variance of W_t^2 will be negligible compared to its expected value. Put another way, the randomness in W_t^2 is negligible compared to its mean, for small t. This suggests that if we take a fine enough partition $\mathcal{P}$ of $[0,t]$, a finite set of points $0 = t_0 < t_1 < \ldots < t_n = t$ with grid mesh $\|\mathcal{P}\| = \max |t_{k+1} - t_k|$ small enough, then writing $D_k := W_{t_{k+1}} - W_{t_k}$ and $\Delta t_k := t_{k+1} - t_k$, we conjecture that $\sum_{k=0}^{n-1} D_k^2$ will "closely resemble"

$$\sum_{k=0}^{n-1} \mathbb{E}[D_k^2] = \sum_{k=0}^{n-1} \Delta t_k = t.$$

This can be made rigorous, as we show below, and the limit of $\sum_{k=0}^{n-1} D_k^2$ as the partition becomes finer is the quadratic variation of Brownian motion over the interval $[0,t]$.

We shall first prove that the quadratic variation of Brownian motion over $[0,t]$ is equal to t in mean square, and then we shall prove that the result holds almost surely (the almost sure convergence proof is not examinable).

Recall that a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables converges *in mean square* (or in $L_2(\Omega, \mathcal{F}, \mathbb{P})$) to a random variable X if $\mathbb{E}[|X_n - X|^2] \to 0$ as $n \to \infty$, and converges to X *almost surely* if $\mathbb{P}\{\omega \in \Omega | X_n(\omega) = X(\omega)\} \to 1$ as $n \to \infty$.

Proof of Theorem 9.10 I: convergence in L_2. Let $\mathcal{P} = \{t_0, t_1, \ldots, t_n\}$ be a partition of $[0, t]$. Set $D_k := W_{t_{k+1}} - W_{t_k}$ and define the *sample quadratic variation*

$$Q_{\mathcal{P}} := \sum_{k=0}^{n-1} D_k^2.$$

Then

$$Q_{\mathcal{P}} - t = \sum_{k=0}^{n-1} [D_k^2 - (t_{k+1} - t_k)].$$

We want to show that $\lim_{\|\mathcal{P}\| \to 0} (Q_{\mathcal{P}} - t) = 0$ in mean square. Consider an individual summand $D_k^2 - (t_{k+1} - t_k)$. This has expectation zero, so

$$\mathbb{E}[Q_{\mathcal{P}} - t] = \mathbb{E} \sum_{k=0}^{n-1} [D_k^2 - (t_{k+1} - t_k)] = 0.$$

Therefore, if we compute $\mathbb{E}[(Q_{\mathcal{P}} - t)^2] = \mathrm{var}(Q_{\mathcal{P}} - t)$ and find it to approach zero as $\|\mathcal{P}\| \to 0$, then we have shown that the quadratic variation of Brownian motion is equal to t in mean square or, equivalently, that $\mathrm{var}(Q_{\mathcal{P}}) \to 0$ as $\|\mathcal{P}\| \to 0$, so that $Q_{\mathcal{P}}$ essentially becomes non-stochastic as $\|\mathcal{P}\| \to 0$.

For $j \neq k$, the terms $D_j^2 - (t_{j+1} - t_j)$ and $D_k^2 - (t_{k+1} - t_k)$ are independent (due to the independent increments property of BM), so

$$\begin{aligned}
\mathrm{var}(Q_{\mathcal{P}} - t) &= \sum_{k=0}^{n-1} \mathrm{var}[D_k^2 - (t_{k+1} - t_k)] \\
&= \sum_{k=0}^{n-1} \mathbb{E}[D_k^4 - 2(t_{k+1} - t_k) D_k^2 + (t_{k+1} - t_k)^2] \\
&= \sum_{k=0}^{n-1} [3(t_{k+1} - t_k)^2 - 2(t_{k+1} - t_k)^2 + (t_{k+1} - t_k)^2] \\
&= 2 \sum_{k=0}^{n-1} (t_{k+1} - t_k)^2 \\
&\leq 2\|\mathcal{P}\| \sum_{k=0}^{n-1} (t_{k+1} - t_k) \\
&= 2\|\mathcal{P}\| t.
\end{aligned}$$

Thus we have

$$\mathbb{E}(Q_{\mathcal{P}} - t) = 0, \quad \mathrm{var}(Q_{\mathcal{P}} - t) \leq 2\|\mathcal{P}\| t.$$

As $\|\mathcal{P}\| \to 0$, $\mathrm{var}(Q_{\mathcal{P}} - t) \to 0$, or $\mathbb{E}[(Q_{\mathcal{P}} - t)^2] \to 0$ as $\|\mathcal{P}\| \to 0$ (i.e. as $n \to \infty$), so

$$Q_{\mathcal{P}} \to t, \quad \text{in } L_2.$$

□

Remark 9.11 (Mean square versus almost sure convergence)*.* We have shown mean square convergence (or L_2 convergence) of $[W]_t$ to t. When such convergence takes place, there is a subsequence of times (so another partititon of $[0,t]$) along which the convergence is *almost sure* (that is, the convergence takes place for all paths except for a set of paths having probability zero). We shall not delve into the subtle differences among different modes of convergence of random variables, but the next proof shows how one can establish almost sure convergence of $[W]_t$ to t.

Proof of Theorem 9.10 II: a.s. convergence.* To show that the convergence is also almost sure, consider the dyadic partition $t_k = kt/2^m, k = 0, 1, \ldots, 2^m$, i.e. we partition $[0,t]$ into 2^m intervals of width $t/2^m$, so that the mesh of the partition approaches zero as $m \to \infty$. Then the sample quadratic variation over $[0,t]$ may be written as

$$Q_m(t) := \sum_{k=0}^{2^m-1} \left(W_{(k+1)t/2^m} - W_{kt/2^m}\right)^2 =: \sum_{k=0}^{2^m-1} (\Delta W_k)^2,$$

where we have written $\Delta W_k = W_{(k+1)t/2^m} - W_{kt/2^m}$. We have $\Delta W_k \sim N(0, t/2^m)$, $\Delta W_k, \Delta W_j$ are independent for $k \neq j$, and hence $(\Delta W_k)^2, (\Delta W_j)^2$ are independent for $k \neq j$.

Recall that for $X \sim N(0, v)$ we have $\mathbb{E}[X^4] = 3v^2$, so that

$$\mathrm{var}[X^2] = \mathbb{E}[X^4] - (\mathbb{E}[X^2])^2 = 3v^2 - v^2 = 2v^2.$$

Therefore, from $\mathbb{E}[\Delta W_k^2] = t/2^m$ we get

$$\mathbb{E}[Q_m(t)] = t,$$

regardless of m. Further, by the independence of the squared increments we have

$$\begin{aligned}
\mathbb{E}[(Q_m(t) - t)^2] &= \mathrm{var}(Q_m(t)) \\
&= \mathrm{var}\left(\sum_{k=0}^{2^m-1} (\Delta W_k)^2\right) \\
&= \sum_{k=0}^{2^m-1} \mathrm{var}(\Delta W_k^2) \\
&= 2^m \cdot 2 \left(\frac{t}{2^m}\right)^2 \\
&= \frac{2t^2}{2^m} \to 0, \quad \text{as } m \to \infty.
\end{aligned}$$

Therefore, since the limit of $Q_m(t)$ as $m \to \infty$ is $[W]_t$. we have established the mean square convergence

$$[W]_t = \lim_{m \to \infty} Q_m(t) \to t, \quad \text{in } L_2.$$

Now we show almost sure convergence using the Chebyshev inequality and the Borel-Cantelli lemmas (see, for instance Grimmett and Stirzaker [5], Section 7.3).[5] By Chebyshev's inequality we have, for $a > 0$,

$$\mathbb{P}\{|Q_m(t) - t| > a\} \leq \frac{1}{a^2}\mathbb{E}[(Q_m(t) - t)^2] = \frac{2t^2}{a^2 2^m}.$$

So

$$\mathbb{P}\{|Q_m(t) - t| > 1/m\} \leq m^2\mathbb{E}[(Q_m(t) - t)^2] = \frac{2t^2 m^2}{2^m}.$$

Write $A_m = \{|Q_m(t) - t| > 1/m\}$, and consider the sequence of events $(A_m)_{m=1}^{\infty}$. Then $\sum_{m=1}^{\infty} \mathbb{P}(A_m) < \infty$, so by the Borel-Cantelli lemmas, the event that infinitely many of the A_m occur has probability given by

$$\mathbb{P}\left(\limsup_{m\to\infty} A_m\right) = \mathbb{P}\left(\bigcap_{m=1}^{\infty}\bigcup_{k=m}^{\infty} A_k\right) = 0.$$

In other words, $|Q_m(t) - t| \leq 1/m$ for large m, almost surely, or

$$[W]_t = \lim_{m\to\infty} Q_m(t) \to t, \quad \text{almost surely.}$$

$\square$

[5] Chebyshev's inequality follows from the following result, which is Theorem 7.3.1 in [5].

Theorem 9.12. *Let $h : \mathbb{R} \to [0, \infty)$ be a non-negative function, Then*

$$\mathbb{P}(h(X) \geq a) \leq \frac{\mathbb{E}[h(X)]}{a}, \quad a > 0.$$

Proof. Let $A := \{h(X) \geq a\}$. Then $h(X) \geq a\mathbb{1}_A$. Taking expectations gives the result.

$\square$

Setting $h(x) = |x|$ gives Markov's inequality. Taking $h(x) = x^2$ gives Chebyshev's inequality: $\mathbb{P}(|X| \geq a) \leq \mathbb{E}[X^2]/a^2$.

The Borel-Cantelli lemmas (Theorem 7.3.10 in [5]) state:

Theorem 9.13 (Borel-Cantelli lemmas)**.** *Let $A_1, A_2, \ldots$ be an infinite sequence of events from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let A be the event that infinitely many of the A_n occur (or $\{A_n \text{ infinitely often}\} = \{A_n \text{ i.o.}\}$, given by*

$$A := \{A_n \text{ i.o.}\} = \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k.$$

Then:

1. $\mathbb{P}(A) = 0$ *if* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$,
2. $\mathbb{P}(A) = 1$ *if* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ *and* $A_1, A_2 \ldots$ *are independent events.*

9.6 Path length*

Given a continuous function $f : [0,t] \to \mathbb{R}$ its total variation over $[0,t]$ is, over any partition $\mathcal{P} = \{0 = t_0 \le t_1 \le \ldots \le t_n = t\}$ of $[0,t]$,

$$TV(f) \equiv [f,f]_t^{(1)} = \lim_{\|\mathcal{P}\|\to 0} \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|.$$

This may be infinite, or some finite number, in which case we say that f has *bounded variation.*

Consider an element of arc length Δs_i along f in the interval $[t_i, t_{i+1}]$. If this interval is small, we have $(\Delta s_i)^2 \approx (\Delta t_i)^2 + (\Delta f_i)^2$, where we have written $\Delta t_i = t_{i+1} - t_i$ and $\Delta f_i = f(t_{i+1}) - f(t_i)$. By the triangle inequality we have

$$|\Delta f_i| \le |\Delta s_i| \le |\Delta f_i| + |\Delta t_i|.$$

Denoting the total arc length (or path length) of f over $[0,t]$ by $s[f]$ we therefore have, in the limit $\|\mathcal{P}\| \to 0$,

$$TV(f) \le s(f) \le TV(f) + t.$$

Therefore,

$$\text{finite path length} \iff TV(f) < \infty.$$

In contrast, the quadratic variation of f over $[0,t]$ is

$$\begin{aligned}
[f]_t &= \lim_{\|\mathcal{P}\|\to 0} \sum_{i=0}^{n-1} |\Delta f_i||\Delta f_i| \\
&\le \lim_{\|\mathcal{P}\|\to 0} \left(\max_{i=0,\ldots,n-1} |\Delta f_i| \right) \lim_{\|\mathcal{P}\|\to 0} \sum_{i=0}^{n-1} |\Delta f_i| \\
&= \lim_{\|\mathcal{P}\|\to 0} \left(\max_{i=0,\ldots,n-1} |\Delta f_i| \right) TV(f).
\end{aligned}$$

For any continuous function, $\lim_{\|\mathcal{P}\|\to 0} (\max_{i=0,\ldots,n-1} |\Delta f_i|) \to 0$[6], so we conclude that

$$TV(f) < \infty \iff [f]_t = 0 \quad \text{for all } t \ge 0.$$

In other words, *paths of Brownian motion $(W_s)_{0\le s\le t}$ over the interval $[0,t]$ have infinite path length.*

Because the total variation of Brownian motion is infinite (i.e. Brownian paths are "very long") one is not able to give meaning to integrals with respect to Brownian motion, $\int_0^t b_s \, \mathrm{d}W_s$, via a path-by-path procedure. Thus we are led to a new type of integral, the *Itô stochastic integral*, which we shall describe shortly.

[6] This is a standard theorem from real analysis, proven from compactness arguments.

Remark 9.14 (Heuristics). If we (formally) write $\mathrm{d}W_t$ for the infinitesimal (corresponding to the infinitesimal time interval $\mathrm{d}t$) increase in W_t, then we have "$\int_0^t \mathrm{d}W_s\,\mathrm{d}W_s = t$", which is often summarised by the formula

$$\mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t.$$

A better way to write this would be

$$\mathrm{d}[W]_t = \mathrm{d}t.$$

Formally, note that if $\mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t$, then in some sense $\mathrm{d}W_t/\,\mathrm{d}t = 1/\sqrt{\mathrm{d}t} \to \infty$ as $\mathrm{d}t \to 0$. In other words, Brownian motion is nowhere differentiable, as we saw earlier.

For the partition $\mathcal{P}$ defined by

$$0 = t_0 < t_1 < \ldots < t_n = t,$$

we defined

$$D_k := W_{t_{k+1}} - W_{t_k}, \quad \Delta t_k := t_{k+1} - t_k, \quad k = 0, 1, \ldots, n-1.$$

We have that

$$\mathbb{E}[D_k^2] = \Delta t_k, \quad \operatorname{var}(D_k^2) = 2(\Delta t_k)^2.$$

It is tempting to argue that, because the variance of D_k^2 is much smaller than its mean, then we have that for small Δt_k, $D_k^2 \approx \Delta t_k$. But this equation has no content: when Δt_k is small, it would be true because both sides are near zero. A better way to capture what we think is going might be to write

$$\frac{D_k^2}{\Delta t_k} \approx 1.$$

But this is never true either. The left hand side is the square of the standard normal random variable

$$Y_k := \frac{D_k}{\sqrt{\Delta t_k}} \sim N(0, 1),$$

whose distribution is the same no matter how small we make Δt_k.

To better understand what is going on, for some large positive integer n, define $t_k := kt/n, k = 0, 1, \ldots, n$, so that $\Delta t_k = t/n$ for all $k = 0, 1, \ldots, n-1$. Then

$$D_k^2 = t\frac{Y_k^2}{n}, \quad k = 0, 1, \ldots, n-1.$$

The random variables $Y_0, Y_1, \ldots, Y_{n-1}$ are i.i.d., so the Law of Large Numbers implies that $\frac{1}{n}\sum_{k=0}^{n-1} Y_k^2$ converges to the common mean $\mathbb{E}[Y_k^2] = 1$ as $n \to \infty$, and hence $\sum_{k=0}^{n-1} D_k^2$ converges to t. Each of the terms D_k^2 in this sum can be quite different from its mean $\Delta t_k = t/n$, but when we sum many terms like this, the differences average out to zero.

So the point is that although we write $\mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t$ frequently, this has no rigorous mathematical meaning unless we consider the integrated relation $[W]_t = t$.

Remark 9.15. For the partition $\mathcal{P}$ defined by

$$0 = t_0 < t_1 < \ldots < t_n = t,$$

we have computed the quadratic variation

$$[W]_t := \lim_{\|\mathcal{P}\| \to 0} \sum_{k=0}^{n-1} D_k^2 = t. \tag{9.5}$$

In addition to this, we can compute the cross variation of W_t with t or the quadratic variation of t, given by

$$\lim_{\|\mathcal{P}\| \to 0} \sum_{k=0}^{n-1} D_k \Delta t_k = 0, \qquad \lim_{\|\mathcal{P}\| \to 0} \sum_{k=0}^{n-1} (\Delta t_k)^2 = 0. \tag{9.6}$$

We know that the second of these limits is zero since t is a differentiable function (see Lemma 9.9, or the argument below). To see that the first limit in (9.6) is zero, observe that

$$|D_k \Delta t_k| \leq \left(\max_{0 \leq j \leq n-1} |D_j| \right) \Delta t_k,$$

and hence

$$\left| \sum_{k=0}^{n-1} D_k \Delta t_k \right| \leq \left(\max_{0 \leq j \leq n-1} |D_j| \right) t,$$

which converges to zero as $\|\mathcal{P}\| \to 0$ since W is continuous.

For the second equality in (9.6) we observe that

$$\sum_{k=0}^{n-1} (\Delta t_k)^2 \leq \|\mathcal{P}\| \sum_{k=0}^{n-1} \Delta t_k = \|\mathcal{P}\| t,$$

which clearly converges to zero as $\|\mathcal{P}\| \to 0$.

Just as we informally write $\mathrm{d}W_t \, \mathrm{d}W_t = \mathrm{d}t$ for (9.5), we capture (9.6) by writing

$$\mathrm{d}W_t \, \mathrm{d}t = 0, \qquad \mathrm{d}t \, \mathrm{d}t = 0.$$

9.7 Other variations of Brownian motion

Consider the *first variation* (or total variation) of BM, denoted by

$$TV(W)_t := \lim_{\|\mathcal{P}\| \to 0} \sum_{k=0}^{n-1} |D_k|.$$

Lemma 9.16. *The first variation of BM is infinite, $TV(W)_t = \infty$, for $t > 0$.*

Proof. We have

$$t = [W]_t = \lim_{\|\mathcal{P}\|\to 0} \sum_{k=0}^{n-1} D_k^2 \le \left(\lim_{\|\mathcal{P}\|\to 0} \max_{0\le j\le n-1} |D_j| \right) TV(W)_t.$$

So $TV(W)_t < \infty$ is equvalent to $[W]_t = 0$, which is false, so the result follows.

□

For the third variation, defined by

$$[W, W]_t^{(3)} := \lim_{\|\mathcal{P}\|\to 0} \sum_{k=0}^{n-1} |D_k|^3,$$

we have:

Lemma 9.17.

$$[W, W]_t^{(3)} = 0, \quad t \ge 0.$$

Proof.

$$[W, W]_t^{(3)} := \lim_{\|\mathcal{P}\|\to 0} \sum_{k=0}^{n-1} |D_k|^3 \le [W]_t \left(\lim_{\|\mathcal{P}\|\to 0} \max_{0\le j\le n-1} |D_j| \right) = 0.$$

□

9.8 Lévy's characterisation of Brownian motion*

BM W is a martingale with continuous paths whose quadratic variation is $[W]_t = t$. In fact, this is a complete characterisation of BM, given in the following Theorem (see Shreve [13], Section 4.6.3 for more details).

Theorem 9.18 (Lévy's theorem, 1-dimensional)**.** *Let M be a martingale relative to a filtration, with $M_0 = 0$, continuous paths, and $[M]_t = t$ for all $t \ge 0$. Then M is a BM.*

10 The Itô integral

We consider how to define an integral with respect to Brownian motion. The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (with $\mathbb{F} = (\mathcal{F})_{t\ge 0}$ the filtration generated by Brownian motion) is given, and always lurks in the background, even when not explicitly mentioned. Recall that Brownian motion, $W_t(\omega) : [0, \infty) \times \Omega \to \mathbb{R}$ has the properties:

1. $W_0 = 0$; (technically, $\mathbb{P}\{\omega : W_0(\omega) = 0\} = 1$);

2. W_t is a continuous function of t;

3. If $0 = t_0 \le t_1 \le \ldots \le t_n = t$, then the increments

   $$W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$$

   are independent, normal, with

   $$\begin{aligned} \mathbb{E}[W_{t_{k+1}} - W_{t_k}] &= 0, \\ \mathbb{E}[(W_{t_{k+1}} - W_{t_k})^2] &= t_{k+1} - t_k, \quad k = 0, 1, \ldots, n-1. \end{aligned}$$

10.1 Construction of the Itô integral

We want to construct the *Itô integral*, which we write as

$$I_t = \int_0^t b_s \, \mathrm{d}W_s, \quad t \geq 0.$$

The integrator is Brownian motion, $(W_t)_{t\geq 0}$, with associated filtration $(\mathcal{F}_t)_{t\geq 0}$ and the following properties:

1. $s \leq t \Rightarrow$ every set in $\mathcal{F}_s$ is also in $\mathcal{F}_t$;
2. W_t is $\mathcal{F}_t$-measurable, $\forall t \geq 0$;
3. for $t \leq t_1 \leq \ldots \leq t_n$, the increments $W_{t_1} - W_t, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent of $\mathcal{F}_t$.

The integrand is a process $b = (b_t)_{t\geq 0}$, where

1. b_t is $\mathcal{F}_t$-measurable $\forall t \geq 0$ (i.e. $(b_t)_{t\geq 0}$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$);
2. b is square-integrable:
$$\mathbb{E}\left[\int_0^t b_s^2 \, \mathrm{d}s\right] < \infty, \quad \forall \quad t \geq 0.$$

Remark 10.1. For a *differentiable* function $f(t)$, we can define

$$\int_0^t b(s) \, \mathrm{d}f(s) = \int_0^t b(s) f'(s) \, \mathrm{d}s.$$

This won't work when the integrator is Brownian motion, because the paths of Brownian motion are not differentiable.

10.2 Itô Integral of an elementary integrand

Let $\mathcal{P} = \{t_0, t_1, \ldots, t_n\}$ be a partition of $[0, t]$, i.e.

$$0 = t_0 \leq t_1 \leq \ldots \leq t_n = t.$$

Assume that b is constant on each interval $[t_k, t_{k+1})$, such that for $t^* \in [t_k, t_{k+1})$, $b_{t^*} = b_{t_k}$. We call such a process b an *elementary process*, or a *simple process*. The Itô integral I_t of such a process is defined by

$$I_t := \sum_{k=0}^{n-1} b_{t_k}(W_{t_{k+1}} - W_{t_k}), \quad t \geq 0.$$

Remark 10.2 (Interpretation as gains from trading). We can interpret the functions b and W as follows:

- Think of W_t as the *price per share* of an asset at time t.
- Think of $t_0, t_1, \dots, t_n$ as the *trading dates* for the asset.
- Think of b_{t_k} as the *number of shares* of the asset held in the interval $[t_k, t_{k+1})$, i.e. acquired at trading date t_k and held until trading date t_{k+1} (so the process π, defined such that $\pi_{t_{k+1}} = b_{t_k}$, is a predictable process).

Then the Ito integral I_t can be interpreted as the *gain from trading* at time t.

Definition 10.3 (Itô integral of elementary process)**.** If $t_k \leq t \leq t_{k+1}$, then the Itô integral $I_t = \int_0^t b_s \,\mathrm{d}W_s$ of the elementary process b is defined by

$$I_t := \int_0^t b_s \,\mathrm{d}W_s := \sum_{j=0}^{k-1} b_{t_j}(W_{t_{j+1}} - W_{t_j}) + b_{t_k}(W_t - W_{t_k}), \quad t \geq 0.$$

10.3 Properties of the Itô integral of an elementary process

Adaptedness For each $t \geq 0$, I_t is $\mathcal{F}_t$-measurable;

Linearity With

$$I_t = \int_0^t b_s \,\mathrm{d}W_s, \quad J_t = \int_0^t a_s \,\mathrm{d}W_s,$$

then for $\alpha, \beta \in \mathbb{R}$,

$$\alpha I_t + \beta J_t = \int_0^t (\alpha b_s + \beta a_s) \,\mathrm{d}W_s.$$

Martingale property $(I_t)_{t\geq 0}$ is a martingale. Let us prove this for the case of an integrand which is an elementary process.

Theorem 10.4 (Martingale property)**.** *The process $I = (I_t)_{t\geq 0}$ defined by*

$$I_t := \sum_{j=0}^{k-1} b_{t_j}(W_{t_{j+1}} - W_{t_j}) + b_{t_k}(W_t - W_{t_k}),$$

is a $(\mathbb{P}, \mathbb{F})$-martingale.

Proof. Let $0 \leq s \leq t$ be given. We treat the more difficult case that s and t are in different subintervals, i.e. there are partition points t_ℓ and t_k such that $s \in [t_\ell, t_{\ell+1}]$ and $t \in [t_k, t_{k+1}]$.

Write

$$\begin{aligned} I_t &= \sum_{j=0}^{k-1} b_{t_j}(W_{t_{j+1}} - W_{t_j}) + b_{t_k}(W_t - W_{t_k}) \\ &= \sum_{j=0}^{\ell-1} b_{t_j}(W_{t_{j+1}} - W_{t_j}) \\ &+ b_{t_\ell}(W_{t_{\ell+1}} - W_{t_\ell}) + \sum_{j=\ell+1}^{k-1} b_{t_j}(W_{t_{j+1}} - W_{t_j}) + b_{t_k}(W_t - W_{t_k}). \end{aligned} \tag{10.1}$$

Compute conditional expectations. For $0 \le s \le t$, we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=0}^{\ell-1} b_{t_j}(W_{t_{j+1}} - W_{t_j}) \middle| \mathcal{F}_s\right] &= \sum_{j=0}^{\ell-1} b_{t_j}(W_{t_{j+1}} - W_{t_j}), \\
\mathbb{E}\left[b_{t_\ell}(W_{t_{\ell+1}} - W_{t_\ell})|\mathcal{F}_s\right] &= b_{t_\ell}(\mathbb{E}[W_{t_{\ell+1}}|\mathcal{F}_s] - W_{t_\ell}) = b_{t_\ell}(W_s - W_{t_\ell}).
\end{aligned}$$

These are the conditional expectations of the first first two terms on the RHS of (10.1). They add up to I_s and so contribute this to $\mathbb{E}[I_t|\mathcal{F}_s]$. We show that the third and fourth terms contribute zero:

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=\ell+1}^{k-1} b_{t_j}(W_{t_{j+1}} - W_{t_j}) \middle| \mathcal{F}_s\right] &= \sum_{j=\ell+1}^{k-1} \mathbb{E}\left[\mathbb{E}\left[b_{t_j}(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_j}\right]|\mathcal{F}_s\right] \\
&= \sum_{j=\ell+1}^{k-1} \mathbb{E}\left[b_{t_j}(\mathbb{E}[W_{t_{j+1}}|\mathcal{F}_{t_j}] - W_{t_j})|\mathcal{F}_s\right] = 0,
\end{aligned}$$

and

$$\mathbb{E}\left[b_{t_k}(W_t - W_{t_k})|\mathcal{F}_s\right] = \mathbb{E}\left[b_{t_k}(\mathbb{E}[W_t|\mathcal{F}_{t_k}] - W_{t_k})|\mathcal{F}_s\right] = 0.$$

□

The Itô isometry Because $(I_t)_{t\ge 0}$ is a martingale and $I_0 = 0$ we have $\mathbb{E}[I_t] = 0$ for all $t \ge 0$. It follows that $\mathrm{var}(I_t) = \mathbb{E}[I_t^2]$, a quantity given by the formula in the next theorem.

Theorem 10.5 (Itô isometry)**.** *The Itô integral of the elementary process b, defined by*

$$I_t := \sum_{j=0}^{k-1} b_{t_j}(W_{t_{j+1}} - W_{t_j}) + b_{t_k}(W_t - W_{t_k}), \tag{10.2}$$

satisfies

$$\mathbb{E}[I_t^2] = \mathbb{E}\left[\int_0^t b_s^2\,\mathrm{d}s\right], \quad t \ge 0.$$

Proof. To simplify notation, write $D_j = W_{t_{j+1}} - W_{t_j}$, $j = 0, \dots, k-1$ and $D_k = W_t - W_{t_k}$, so that (10.2) is written as $I_t = \sum_{j=0}^{k} b_{t_j} D_j$. Then

$$I_t^2 = \sum_{j=0}^{k} b_{t_j}^2 D_j^2 + 2 \sum_{0 \le i < j \le k} b_{t_i} b_{t_j} D_i D_j.$$

First we show that the expected value of the cross terms is zero. For $i < j$, the random variable $b_{t_i} b_{t_j} D_i$ is $\mathcal{F}_{t_j}$-measurable, while the Brownian increment D_j is independent of $\mathcal{F}_{t_j}$, so $\mathbb{E}[D_j|\mathcal{F}_{t_j}] = \mathbb{E}[D_j] = 0$. Therefore,

$$\begin{aligned}
\mathbb{E}[b_{t_i} b_{t_j} D_i D_j] &= \mathbb{E}\left[\mathbb{E}[b_{t_i} b_{t_j} D_i D_j|\mathcal{F}_{t_j}]\right] \\
&= \mathbb{E}\left[b_{t_i} b_{t_j} D_i \mathbb{E}[D_j|\mathcal{F}_{t_j}]\right] \\
&= 0.
\end{aligned}$$

Now consider the square terms $b_{t_j}^2 D_j^2$. The random variable $b_{t_j}^2$ is $\mathcal{F}_{t_j}$-measurable, while the squared Brownian increment D_j^2 is independent of $\mathcal{F}_{t_j}$, so $\mathbb{E}[D_j^2|\mathcal{F}_{t_j}] = \mathbb{E}[D_j^2] = t_{j+1} - t_j$, for $j = 0, \ldots, k-1$, and $\mathbb{E}[D_k^2|\mathcal{F}_{t_k}] = \mathbb{E}[D_k^2] = t - t_k$. Therefore,

$$
\begin{aligned}
\mathbb{E}[I_t^2] &= \sum_{j=0}^{k} \mathbb{E}[b_{t_j}^2 D_j^2] \\
&= \sum_{j=0}^{k} \mathbb{E}\left[\mathbb{E}[b_{t_j}^2 D_j^2|\mathcal{F}_{t_j}]\right] \\
&= \sum_{j=0}^{k} \mathbb{E}[b_{t_j}^2 \mathbb{E}[D_j^2|\mathcal{F}_{t_j}]] \\
&= \sum_{j=0}^{k} \mathbb{E}[b_{t_j}^2 \mathbb{E}[D_j^2]] \\
&= \sum_{j=0}^{k-1} \mathbb{E}[b_{t_j}^2 (t_{j+1} - t_j)] + \mathbb{E}[b_{t_k}^2 (t - t_k)].
\end{aligned}
$$

But b_{t_j} is constant on $[t_j, t_{j+1})$, so $b_{t_j}^2 (t_{j+1} - t_j) = \int_{t_j}^{t_{j+1}} b_s^2 \,\mathrm{d}s$ and similarly, $b_{t_k}^2 (t - t_k) = \int_{t_k}^{t} b_s^2 \,\mathrm{d}s$, so

$$
\begin{aligned}
\mathbb{E}[I_t^2] &= \sum_{j=0}^{k-1} \left(\mathbb{E}\left[\int_{t_j}^{t_{j+1}} b_s^2 \,\mathrm{d}s\right]\right) + \mathbb{E}\left[\int_{t_k}^{t} b_s^2 \,\mathrm{d}s\right] \\
&= \mathbb{E}\left[\sum_{j=0}^{k-1} \left(\int_{t_j}^{t_{j+1}} b_s^2 \,\mathrm{d}s\right) + \int_{t_k}^{t} b_s^2 \,\mathrm{d}s\right] \\
&= \mathbb{E}\left[\int_0^t b_s^2 \,\mathrm{d}s\right].
\end{aligned}
$$

□

Quadratic variation of the integral The quadratic variation of the integral, thought of as the quadratic variation process $([I]_t)_{t\geq 0}$ of the integral process $I = (I_t)_{t\geq 0}$. Brownian motion $W_t = \int_0^t 1 \cdot \mathrm{d}W_s$ has quadratic variation $[W]_t = \int_0^t 1^2 \cdot \mathrm{d}[W]_s = \int_0^t 1 \cdot \mathrm{d}t$. We say that Brownian motion *accumulates quadratic variation at the rate of one per unit time.* In the Itô integral $I_t = \int_0^t b_s \,\mathrm{d}W_s$, BM is scaled in a time and path-dependent way (i.e. depending on $(s, \omega) \in [0, t] \times \Omega$) by the integrand b_s. Because increments are squared in the computation of quadratic variation, the QV of BM will be scaled by b_s^2 as it enters the integral. The following theorem gives the precise statement.

Theorem 10.6 (Quadratic variation of the Itô integral)**.** *Let b be a simple process. Then the Itô integral*

$$
I_t = \int_0^t b_s \,\mathrm{d}W_s, \quad t \geq 0,
$$

has quadratic variation process $([I]_t)_{t\geq 0}$ *given by*

$$[I]_t = \int_0^t b_s^2 \,\mathrm{d}s, \quad t \geq 0.$$

We say that the Itô integral accumulates quadratic variation at a rate $b_s^2, s \in [0,t]$ per unit time, and that the quadratic variation accumulated up to time t by the integral is $[I]_t = \int_0^t b_s^2 \,\mathrm{d}s$.

Proof. First compute the quadratic variation accumulated by the integral on one of the subintervals $[t_j, t_{j+1})$ on which $b_s = b_{t_j}, s \in [t_j, t_{j+1})$, is constant. Choose partition points

$$t_j = s_0 < s_1 < \ldots < s_m = t_{j+1},$$

and consider

$$\begin{aligned} \sum_{i=0}^{m-1} (I_{s_{i+1}} - I_{s_i})^2 &= \sum_{i=0}^{m-1} [b_{t_j}(W_{s_{i+1}} - W_{s_i})]^2 \\ &= b_{t_j}^2 \sum_{i=0}^{m-1} (W_{s_{i+1}} - W_{s_i})^2. \end{aligned} \tag{10.3}$$

As $m \to \infty$ and the mesh of the partition, $\max_{i=0,\ldots,m-1}(s_{i+1} - s_i)$ approaches zero, the term $\sum_{i=0}^{m-1} (W_{s_{i+1}} - W_{s_i})^2$ converges to the QV accumulated by BM over $[t_j, t_{j+1})$, which is $t_{j+1} - t_j$. Therefore, the limit of the RHS of (10.3), which is the QV accumulated by the integral over $[t_j, t_{j+1})$, is

$$b_{t_j}^2 (t_{j+1} - t_j) = \int_{t_j}^{t_{j+1}} b_s^2 \,\mathrm{d}s,$$

where we have use the fact that b_s is constant for $s \in [t_j, t_{j+1})$. Similarly, the QV accumulated by the integral over $[t_k, t]$ is $\int_{t_k}^t b_s^2 \,\mathrm{d}s$. Adding up all these contributions proves the theorem. □

Informally, we establish the theorem in differential form via

$$\mathrm{d}I_t = b_t \,\mathrm{d}W_t \Longrightarrow \mathrm{d}[I]_t = \mathrm{d}I_t \,\mathrm{d}I_t = b_t^2 \,\mathrm{d}W_t \,\mathrm{d}W_t = b_t^2 \,\mathrm{d}[W]_t = b_t^2 \,\mathrm{d}t,$$

just as we wrote $\mathrm{d}[W]_t = \mathrm{d}W_t \,\mathrm{d}W_t = \mathrm{d}t$ earlier. In fact, one can do a lot of the calculations in Itô calculus simply by applying the informal multiplication rules:

$$\mathrm{d}W_t \,\mathrm{d}W_t = \mathrm{d}t, \quad \mathrm{d}W_t \,\mathrm{d}t = \mathrm{d}t \,\mathrm{d}t = 0.$$

Remark 10.7. Note the contrast between Theorems 10.5 and 10.6. The QV $[I]_t$ is computed path-by-path, so the result can depend on the path, and so in principle is random. The variance of the integral is precisely the expectation of the QV, as given by the Itô isometry (i.e. it is is an average over all possible paths of the QV), and so is non-random.

10.4 Itô Integral of a general integrand*

Fix $t > 0$. Let b be a process (not necessarily an elementary process) such that

- b_s is $\mathcal{F}_s$-measurable, $\forall s \in [0,t]$;
- $\mathbb{E}\left[\int_0^t b_s^2 \,\mathrm{d}s\right] < \infty$.

We then have the following result.

Theorem 10.8. *There is a sequence of elementary processes $(b^{(n)})_{n=1}^\infty$ such that*

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^t |b_s^{(n)} - b_s|^2 \,\mathrm{d}s\right] = 0.$$

Proof. See [13], Section 4.3, [11], Section 3.1, or [9], Section 3.2 and Problem 3.2.5 in [9]. □

We have shown how to define

$$I_t^{(n)} = \int_0^t b_s^{(n)} \,\mathrm{d}W_s,$$

for every $n \in \mathbb{N}$. We now define the general Itô integral by

$$\int_0^t b_s \,\mathrm{d}W_s := \lim_{n\to\infty} \int_0^t b_s^{(n)} \,\mathrm{d}W_s.$$

The only difficulty with this approach is that we need to make sure the above limit exists. Suppose m and n are large positive integers. Then

$$\begin{aligned}
\mathbb{E}[|I_t^{(n)} - I_t^{(m)}|^2] &= \operatorname{var}(I_t^{(n)} - I_t^{(m)}) \\
&= \mathbb{E}\left[\left(\int_0^t (b_s^{(n)} - b_s^{(m)}) \,\mathrm{d}W_s\right)^2\right] \\
\text{(Itô isometry)} &= \mathbb{E}\left[\int_0^t (b_s^{(n)} - b_s^{(m)})^2 \,\mathrm{d}s\right] \\
\text{(triangle inequality)} &\le \mathbb{E}\left[\int_0^t (|b_s^{(n)} - b_s| + |b_s - b_s^{(m)}|)^2 \,\mathrm{d}s\right] \\
((a+b)^2 \le 2(a^2+b^2)) &\le 2\mathbb{E}\left[\int_0^t |b_s^{(n)} - b_s|^2 \,\mathrm{d}s\right] + 2\mathbb{E}\left[\int_0^t |b_s - b_s^{(m)}|^2 \,\mathrm{d}s\right],
\end{aligned}$$

which approaches zero as $m, n \to \infty$, by Theorem 10.8. This guarantees that the sequence $(I_t^{(n)})_{n=1}^\infty$ is a Cauchy sequence in $L_2(\Omega, \mathcal{F}, \mathbb{P})$ and so has a limit.

10.5 Properties of the general Itô integral

The general Itô integral is

$$I_t = \int_0^t b_s \, \mathrm{d}W_s,$$

where b is any adapted, square-integrable process. Its properties are inherited from the properties of Itô integrals of simple processes and are summarised below.

Adaptedness For each $t \geq 0$, I_t is $\mathcal{F}_t$-measurable;

Linearity If

$$I_t = \int_0^t b_s \, \mathrm{d}W_s, \quad J_t = \int_0^t a_s \, \mathrm{d}W_s,$$

then for $\alpha, \beta \in \mathbb{R}$,

$$\alpha I_t + \beta J_t = \int_0^t (\alpha b_s + \beta a_s) \, \mathrm{d}W_s.$$

Martingale property $(I_t)_{t \geq 0}$ is a martingale.

In fact, we have the converse result, known as the martingale representation theorem (which we do not prove, see [11] for example).

Itô and martingale representation theorems for Brownian motion*

Theorem 10.9 (Itô representation theorem for Brownian motion). *Let $(W_t)_{t\geq 0}$ be a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, with $(\mathcal{F}_t)_{t\geq 0}$ the natural filtration $\mathcal{F}_t = \sigma(W_s, 0 \leq s \leq t)$. Suppose that $X \in L_2(\Omega, \mathcal{F}_t, \mathbb{P})$ (i.e. X is $\mathcal{F}_t$-measurable and $\mathbb{E}[X^2] < \infty$). Then there exists an adapted process b such that $\mathbb{E}\left[\int_0^t b_s^2 \, \mathrm{d}s\right] < \infty, t \geq 0$ and*

$$X = \mathbb{E}[X] + \int_0^t b_s \, \mathrm{d}W_s.$$

Theorem 10.10 (Martingale representation theorem for Brownian motion). *Let $(W_t)_{t\geq 0}$ be a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, with $(\mathcal{F}_t)_{t\geq 0}$ the natural filtration $\mathcal{F}_t = \sigma(W_s, 0 \leq s \leq t)$. Suppose that the process $M = (M_t)_{t\geq 0}$ is a square-integrable martingale with respect to this filtration, written $M \in \mathcal{M}_2$ (that is, $M_t \in L_2(\Omega, \mathcal{F}_t, \mathbb{P})$ for all $t \geq 0$, or $\mathbb{E}[M_t^2] < \infty$, for all $t \geq 0$). Then there exists an adapted process b such that $\mathbb{E}\left[\int_0^t b_s^2 \, \mathrm{d}s\right] < \infty, t \geq 0$ and*

$$M_t = M_0 + \int_0^t b_s \, \mathrm{d}W_s.$$

The Itô isometry The variance of the Itô integral is $\mathrm{var}(I_t) = \mathbb{E}[I_t^2]$ given by

$$\mathbb{E}[I_t^2] = \mathbb{E}\left[\int_0^t b_s^2 \, \mathrm{d}s\right].$$

Continuity I_t is a continuous function of the upper limit of integration t.

Quadratic variation The Itô integral

$$I_t = \int_0^t b_s \,\mathrm{d}W_s, t \geq 0,$$

has quadratic variation process $([I]_t)_{t\geq 0}$ given by

$$[I]_t = \int_0^t b_s^2 \,\mathrm{d}s.$$

Example 10.11. Consider the Itô integral

$$I_t = \int_0^t W_s \,\mathrm{d}W_s.$$

We approximate the integrand by an elementary process $b_s^{(n)}$, $s \in [0,t]$, in the following way. Partition the interval $[0,t]$ into n time intervals δt, so that $t = n\delta t$, and

$$0 = t_0 < t_1 = \delta t = \frac{t}{n} < \ldots < t_k = k\delta t = \frac{kt}{n} < \ldots < t_n = t,$$

and define $b_s^{(n)}$ by

$$b_s^{(n)} = W_{t_k} = W_{kt/n}, \quad \text{if } \tfrac{kt}{n} \leq s < \tfrac{(k+1)t}{n}, \quad k = 0,\ldots,n-1.$$

Then by definition

$$I_t = \int_0^t W_s \,\mathrm{d}W_s = \lim_{n\to\infty} \sum_{k=0}^{n-1} W_{kt/n} \left(W_{(k+1)t/n} - W_{kt/n}\right).$$

To simplify notation, write $W_k \equiv W_{kt/n}$ so that

$$\int_0^t W_s \,\mathrm{d}W_s = \lim_{n\to\infty} \sum_{k=0}^{n-1} W_k(W_{k+1} - W_k).$$

Then we note that

$$\begin{aligned} W_{k+1}^2 - W_k^2 &= (W_{k+1} - W_k)^2 + 2W_k W_{k+1} - 2W_k^2 \\ &= (W_{k+1} - W_k)^2 + 2W_k(W_{k+1} - W_k), \end{aligned}$$

so that

$$\begin{aligned} \sum_{k=0}^{n-1} W_k(W_{k+1} - W_k) &= \frac{1}{2}\left[\sum_{k=0}^{n-1}(W_{k+1}^2 - W_k^2) - \sum_{k=0}^{n-1}(W_{k+1} - W_k)^2\right] \\ &= \frac{1}{2}\left(W_n^2 - \sum_{k=0}^{n-1}(W_{k+1} - W_k)^2\right). \quad (W_0 = 0) \end{aligned}$$

Now we let $n \to \infty$ and use the definition of quadratic variation to get

$$\int_0^t W_s \, \mathrm{d}W_s = \frac{1}{2}(W_t^2 - [W]_t) = \frac{1}{2}(W_t^2 - t).$$

Remark 10.12 (Reason for the $\frac{1}{2}t$ term). If f is a differentiable function with $f(0) = 0$, then

$$\int_0^t f(s) \, \mathrm{d}f(s) = \int_0^t f(s) f'(s) \, \mathrm{d}s = \frac{1}{2} f^2(s)|_0^t = \frac{1}{2} f^2(t).$$

In contrast, for Brownian motion, we have

$$\int_0^t W_s \, \mathrm{d}W_s = \frac{1}{2}(W_t^2 - t).$$

The extra term $\frac{1}{2}t$ comes from the nonzero quadratic variation of Brownian motion. It has to be there, because $\mathbb{E}[\int_0^t W_s \, \mathrm{d}W_s] = 0$ (the Itô integral is a martingale), but $\mathbb{E}[\frac{1}{2}W_t^2] = \frac{1}{2}t$. Note that this remark is equivalent to our initial characterisation of Brownian motion in Remark 9.6.

11 The Itô formula

11.1 Itô's formula for one Brownian motion

We want a rule to "differentiate" expressions of the form $f(W_t)$. If W_t were differentiable then the ordinary chain rule would give

$$\frac{\mathrm{d}}{\mathrm{d}t} f(W_t) = f'(W_t) W_t',$$

which could be written in differential notation as

$$\mathrm{d}f(W_t) = f'(W_t) W_t' \, \mathrm{d}t = f'(W_t) \, \mathrm{d}W_t.$$

However, W_t is not differentiable, and in particular has nonzero quadratic variation, so the correct formula has an extra term, namely,

$$\mathrm{d}f(W_t) = f'(W_t) W_t' \, \mathrm{d}t = f'(W_t) \, \mathrm{d}W_t + \frac{1}{2} f''(W_t) \, \mathrm{d}[W]_t,$$

with the understanding that $\mathrm{d}[W]_t = \mathrm{d}t$. This is a version of *Itô's formula in differential form.* Integrating this, we obtain a version of *Ito's formula in integral form.*

Theorem 11.1 (Itô formula for one BM)**.** *If $f(x)$ is a $C^2(\mathbb{R})$ function and $t \geq 0$, then*

$$f(W_t) - f(W_0) = \int_0^t f'(W_s) \, \mathrm{d}W_s + \frac{1}{2} \int_0^t f''(W_s) \, \mathrm{d}[W]_s. \tag{11.1}$$

Remark 11.2 (Differential versus integral forms). The mathematically meaningful form of Itô's formula is its integral form, because we have solid definitions for the integrals appearing on the RHS of (11.1). For pencil and paper computations, the more convenient form is the differential form.

Proof of Theorem 11.1. Fix $t > 0$ and let $\mathcal{P} = \{t_0, t_1, \dots, t_n\}$ be a partition of $[0,t]$. By Taylor's theorem we have

$$\begin{aligned} f(W_t) - f(W_0) &= \sum_{k=0}^{n-1}[f(W_{t_{k+1}}) - f(W_{t_k})] \\ &= \sum_{k=0}^{n-1}\left[f'(W_{t_k})(W_{t_{k+1}} - W_{t_k}) + \frac{1}{2}f''(W_{t_k})(W_{t_{k+1}} - W_{t_k})^2 + \cdots\right] \\ &\overset{\|\mathcal{P}\|\to 0}{\longrightarrow} \int_0^t f'(W_s)\,\mathrm{d}W_s + \frac{1}{2}\int_0^t f''(W_s)\,\mathrm{d}[W]_s, \end{aligned}$$

with higher order terms disappearing (since the third variatiion of BM is zero by Lemma 9.17 and $\sum_{k=0}^{n-1} D_k^3 \leq \sum_{k=0}^{n-1} |D_k|^3$, where $D_k := W_{t_{k+1}} - W_{t_k}$) and the last summation converging to a Riemann integral as it becomes the quadratic variation of an Itô integral, i.e. for the Itô integral

$$I_t = \int_0^t b_s\,\mathrm{d}W_s = \lim_{\|\mathcal{P}\|\to 0}\sum_{k=0}^{n-1} b_{t_k}(W_{t_{k+1}} - W_{t_k}),$$

we have

$$\begin{aligned} \int_0^t b_s^2\,\mathrm{d}s = [I]_t &= \lim_{\|\mathcal{P}\|\to 0}\sum_{k=0}^{n-1}(I_{t_{k+1}} - I_{t_k})^2 \\ &= \lim_{\|\mathcal{P}\|\to 0}\sum_{k=0}^{n-1} b_{t_k}^2(W_{t_{k+1}} - W_{t_k})^2. \end{aligned}$$

□

A heuristic derivation would simply state that, by Taylor's theorem

$$df(W_t) = f'(W_t)\,\mathrm{d}W_t + \frac{1}{2}f''(W_t)\,\mathrm{d}t,$$

where we have used $\mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t$ in the last term on the RHS, and higher order terms are neglected.

Corollary 11.3 (Itô formula for function of time and one Brownian motion)**.** *If $S_t = f(t, W_t)$ for some $C^{1,2}(\mathbb{R}^+ \times \mathbb{R})$ function $f(t,x)$, then*

$$\mathrm{d}S_t = \mathrm{d}f(t, W_t) = f_t(t, W_t)\,\mathrm{d}t + f_x(t, W_t)\,\mathrm{d}W_t + \frac{1}{2}f_{xx}(t, W_t\,\mathrm{d}[W]_t,$$

and higher order terms do not contribute, since we have shown earlier (Remark 9.15) that we have the informal rules $\mathrm{d}W_t\,\mathrm{d}t = 0$ and $\mathrm{d}t\,\mathrm{d}t = 0$.

Definition 11.4 (Geometric Brownian motion)**.** Geometric Brownian motion is the process $S = (S_t)_{t\geq 0}$ given by

$$S_t = S_0 \exp\left[\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right) t\right],$$

where μ and $\sigma > 0$ are constant, and the parameter σ is called the *volatility* of the process S.

Define

$$f(t,x) = S_0 \exp\left(\sigma x + (\mu - \frac{1}{2}\sigma^2)t\right),$$

so that $S_t = f(t, W_t)$ and

$$f_t(t,x) = \left(\mu - \frac{1}{2}\sigma^2\right) f(t,x), \quad f_x(t,x) = \sigma f(t,x), \quad f_{xx}(t,x) = \sigma^2 f(t,x),$$

with the subscripts denoting partial derivatives. Then by Itô's formula

$$\begin{aligned}
\mathrm{d}S_t &= \mathrm{d}f(t,W_t) \\
&= f_t(t,W_t)\,\mathrm{d}t + f_x(t,W_t)\,\mathrm{d}W_t + \frac{1}{2} f_{xx}(t,W_t)\,\mathrm{d}t \\
&= \left(\mu - \frac{1}{2}\sigma^2\right) f(t,W_t)\,\mathrm{d}t + \sigma f(t,W_t)\,\mathrm{d}W_t + \frac{1}{2}\sigma^2 f(t,W_t)\,\mathrm{d}t \\
&= \left(\mu - \frac{1}{2}\sigma^2\right) S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t + \frac{1}{2}\sigma^2 S_t\,\mathrm{d}t \\
&= \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t,
\end{aligned}$$

which is geometric Brownian motion in differential form, Geometric Brownian motion in integral form may be written as

$$S_t = S_0 + \int_0^t \mu S_s\,\mathrm{d}s + \int_0^t \sigma S_s\,\mathrm{d}W_s.$$

Quadratic variation of geometric Brownian motion In the integral form of geometric Brownian motion,

$$S_t = S_0 + \int_0^t \mu S_s\,\mathrm{d}s + \int_0^t \sigma S_s\,\mathrm{d}W_s,$$

the Riemann integral

$$F(t) = \int_0^t \mu S_s\,\mathrm{d}s$$

is differentiable with $F'(t) = \mu S_t$. This term has zero quadratic variation. The Itô integral

$$G(t) = \int_0^t \sigma S_s\,\mathrm{d}W_s$$

is not differentiable. It has quadratic variation

$$[G]_t = \int_0^t \sigma^2 S_s^2 \, \mathrm{d}s.$$

Thus the quadratic variation of S is given by the quadratic variation of G, i.e.

$$[S]_t = [G]_t = \int_0^t \sigma^2 S_s^2 \, \mathrm{d}s.$$

In differential notation we write

$$\mathrm{d}[S]_t = \mathrm{d}S_t \, \mathrm{d}S_t = \sigma^2 S_t^2 \, \mathrm{d}t,$$

which follow from the following informal multiplication rules involving the differentials $\mathrm{d}t$ and $\mathrm{d}W_t$:

$$\mathrm{d}[W]_t = \mathrm{d}W_t \, \mathrm{d}W_t = \mathrm{d}t, \quad \mathrm{d}W_t . \mathrm{d}t = \mathrm{d}t . \mathrm{d}W_t = \mathrm{d}t . \mathrm{d}t = 0.$$

Remark 11.5. Note that

$$\int_0^t \frac{\mathrm{d}[S]_s}{S_s^2} = \int_0^t \sigma^2 \, \mathrm{d}s = \sigma^2 t,$$

indicating that for geometric Brownian motion, the quadratic variation, when scaled by the square of the stock price process, is a measure of the volatility of the process S.

11.2 Itô's formula for Itô processes

Definition 11.6 (Itô process)**.** Let $(W_t, \mathcal{F}_t)_{t \geq 0}$ be a standard Brownian motion. An Itô process is a stochastic process of the form

$$X_t = X_0 + \int_0^t a_s \, \mathrm{d}s + \int_0^t b_s \, \mathrm{d}W_s, \quad t \geq 0, \tag{11.2}$$

where X_0 is non-random and a, b are adapted stochastic processes satisfying $\int_0^t |a_s| \, \mathrm{d}s < \infty$ and $\mathbb{E}\left[\int_0^t b_s^2 \, \mathrm{d}s\right] < \infty$.

In differential form we write (11.2) as

$$\mathrm{d}X_t = a_t \, \mathrm{d}t + b_t \, \mathrm{d}W_t.$$

Lemma 11.7 (Quadratic variation of an Itô process)**.** *The quadratic variation of the Itô process (11.2) is the process* $([X]_t)_{t \geq 0}$ *given by*

$$[X]_t = \int_0^t b_s^2 \, \mathrm{d}s, \quad t \geq 0.$$

Proof. This is immediate from the fact that the quadratic variation of $\int_0^t a_s \, \mathrm{d}s$ is zero. □

Definition 11.8 (Integral with respect to an Itô process)**.** Let $(X_t)_{t\geq 0}$ be the Itô process (11.2) and let $(\theta_t)_{t\geq 0}$ be an adapted process satisfying

$$\mathbb{E}\left[\int_0^t \theta_s^2 b_s^2 \,\mathrm{d}s\right] < \infty, \quad \int_0^t |\theta_s a_s| ds < \infty,$$

for every $t \geq 0$. The Itô integral of θ with respect to X is the process $J = (J_t)_{t\geq 0}$ defined by

$$J_t := \int_0^t \theta_s \,\mathrm{d}X_s := \int_0^t \theta_s b_s \,\mathrm{d}W_s + \int_0^t \theta_s a_s \,\mathrm{d}s.$$

Theorem 11.9 (Itô formula for Itô processes)**.** *Let $(X_t)_{t\geq 0}$ be the Itô process (11.2) and let $f(t,x) \in C^{1,2}([0,\infty)\times\mathbb{R})$. Then, for every $t > 0$,*

$$\begin{aligned}
f(t,X_t) &= f(0,X_0) + \int_0^t f_t(s,X_s)\,\mathrm{d}s + \int_0^t f_x(s,X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f_{xx}(s,X_s)\,\mathrm{d}[X]_s \\
&= f(0,X_0) + \int_0^t \left(f_t(s,X_s) + a_s f_x(s,X_s) + \frac{1}{2}b_s^2 f_{xx}(s,X_s)\right)\mathrm{d}s \\
&+ \int_0^t b_s f_x(s,X_s)\,\mathrm{d}W_s.
\end{aligned}$$

Proof. As for the Itô formula with respect to BM, and use the fact that $[X]_t = [I]_t = \int_0^t b_s^2 \,\mathrm{d}s$. □

It is usually easier to remember and use this theorem in the differential form

$$\mathrm{d}f(t,X_t) = f_t(t,X_t)\,\mathrm{d}t + f_x(t,X_t)\,\mathrm{d}X_t + \frac{1}{2}f_{xx}(t,X_t)\,\mathrm{d}[X]_t,$$

where $\mathrm{d}[X]_t = \mathrm{d}X_t\,\mathrm{d}X_t$ is computed according to the rules

$$\mathrm{d}t\,\mathrm{d}t = \mathrm{d}t\,\mathrm{d}W_t = \mathrm{d}W_t\,\mathrm{d}t = 0, \quad \mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t.$$

Example 11.10 (Generalised geometric Brownian motion). Define the Itô process

$$X_t = \int_0^t \sigma_s \,\mathrm{d}W_s + \int_0^t \left(\mu_s - \frac{1}{2}\sigma_s^2\right)\mathrm{d}s, \quad t \geq 0,$$

where μ, σ are adapted processes. Then

$$\begin{aligned}
\mathrm{d}X_t &= \sigma_t \,\mathrm{d}W_t + \left(\mu_t - \frac{1}{2}\sigma_t^2\right)\mathrm{d}t, \\
d[X]_t &= \sigma_t^2 d[W]_t = \sigma_t^2 \,\mathrm{d}t.
\end{aligned}$$

A common model for an asset price process $S = (S_t)_{t\geq 0}$ is given by

$$S_t = S_0 e^{X_t},$$

with $S_0 > 0$ non-random, which is called a generalised geometric Brownian motion. We write $S_t = f(X_t)$ where $f(x) = S_0 e^x$. The Itô formula gives

$$\mathrm{d}S_t = \mu_t S_t \,\mathrm{d}t + \sigma_t S_t \,\mathrm{d}W_t.$$

Applying the Itô formula to the function $g(t, S_t) = \log S_t$, we find that

$$\mathrm{d}(\log S_t) = \mathrm{d}X_t = \sigma_t \,\mathrm{d}W_t + \left(\mu_t - \frac{1}{2}\sigma_t^2\right) \mathrm{d}t.$$

11.3 Stochastic differential equations

Given an Itô process $X := (X_t)_{t\geq 0}$ satisfying

$$X_t = x + \int_0^t \mu_s \,\mathrm{d}s + \int_0^t \sigma_s \,\mathrm{d}W_s,$$

which we usually write in differential form

$$\mathrm{d}X_t = \mu_t \,\mathrm{d}t + \sigma_t \,\mathrm{d}W_t, \tag{11.3}$$

then given a function $f \in C^{1,2}([0,\infty) \times \mathbb{R})$ (i.e. $f = f(t,x)$ for $t \in [0,\infty), x \in \mathbb{R}$, $f : [0,\infty) \times \mathbb{R} \to \mathbb{R}$, differentiable at least once with respect to t and at least twice with respect to x), the process $(Y_t)_{t\geq 0}$ defined by $Y_t := f(t, X_t)$ has differential given by

$$\mathrm{d}Y_t \equiv \mathrm{d}f(t,X_t) = f_t(t,X_t)\,\mathrm{d}t + f_x(t,X_t)\,\mathrm{d}X_t + \frac{1}{2}f_{xx}(t,X_t)\,\mathrm{d}[X]_t,$$

where $\mathrm{d}[X]_t = \mathrm{d}X_t\,\mathrm{d}X_t$ is computed according to the rules

$$\mathrm{d}t\,\mathrm{d}t = \mathrm{d}t\,\mathrm{d}W_t = \mathrm{d}W_t\,\mathrm{d}t = 0, \quad \mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t.$$

In integral form Y_t is given by

$$Y_t = Y_0 + \int_0^t \left(f_t(s,X_s) + \mu_s f_x(s,X_s) + \frac{1}{2}\sigma_s^2 f_{xx}(s,X_s)\right) \mathrm{d}s + \int_0^t \sigma_s f_x(s,X_s)\,\mathrm{d}W_s.$$

11.3.1 Markovian diffusions

If, in (11.3), we have $\mu_t = \mu(t,X_t), \sigma_t = \sigma(t,X_t)$ for well-behaved (see precise conditions later) functions $\mu(t,x), \sigma(t,x)$, so that

$$dX_t = \mu(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}W_t,$$

which is called a *stochastic differential equation* (SDE) for X, then the process X is Markovian:

$$\mathbb{E}[h(X_T)|\mathcal{F}_t] = \mathbb{E}[h(X_T)|X_t], \quad 0 \leq t \leq T,$$

and the integral equation for Y may be written

$$Y_t = Y_0 + \int_0^t (f_t(s,X_s) + \mathcal{A}f(s,X_s))\,\mathrm{d}s + \int_0^t \sigma(s,X_s) f_x(s,X_s)\,\mathrm{d}W_s,$$

where $\mathcal{A}$ is called the *generator* of the *diffusion* X, and is defined by

$$\mathcal{A}f(t,x) := \mu(t,x)f_x(t,x) + \frac{1}{2}\sigma^2(t,x)f_{xx}(t,x).$$

11.3.2 Solutions to stochastic differential equations

We ask whether there exists a well-defined process X satisfying the stochastic differential equation (SDE)

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \tag{11.4}$$

or, more precisely, whether there exists a process X satisfying $X_0 = x$ and

$$X_t = x + \int_0^t \mu(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s, \quad t \geq 0.$$

The basic existence result is as follows. Suppose there is a constant K such that for all x, y, t we have

$$|\mu(t,x) - \mu(t,y)| \leq K|x-y|, \quad |\sigma(t,x) - \sigma(t,y)| \leq K|x-y| \quad |\mu(t,x)| + |\sigma(t,x)| \leq K(1+|x|).$$

(The first two conditions are *Lipschitz continuity* in x.) Then the SDE (11.4) has a unique, adapted, continuous Markovian solution, and there exists a constant C such that

$$\mathbb{E}[|X_t|^2] \leq Ce^{Ct}(1+|x|^2).$$

Example 11.11 (Exponential martingales). Let θ be a process adapted to the filtration of the Brownian motion W. Define the process $Z = (Z_t)_{0 \leq t \leq T}$ by

$$Z_t = \exp\left(-\int_0^t \theta_s\,\mathrm{d}W_s - \frac{1}{2}\int_0^t \theta_s^2 d[W]_s\right).$$

In Problem Sheet 2 we show via the Itô formula that

$$\mathrm{d}Z_t = -\theta_t Z_t\,\mathrm{d}W_t,$$

so Z is the Itô process given by

$$Z_t = 1 - \int_0^t \theta_s Z_s\,\mathrm{d}W_s, \quad t \geq 0,$$

and Z is a martingale provided that $\mathbb{E}\left[\int_0^T \theta_t^2 Z_t^2\,\mathrm{d}t\right] < \infty$.

Remark 11.12 (Novikov condition). A sufficient condition for Z to be a martingale is the *Novikov condition*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \theta_t^2\,\mathrm{d}t\right)\right] < \infty.$$

11.4 Multidimensional Brownian motion

Definition 11.13 (d-dimensional Brownian Motion)**.** A d-dimensional Brownian motion is a process

$$W_t = (W_t^1, \ldots, W_t^d)$$

with the following properties:

- Each W_t^i $(i = 1, \ldots, d)$ is a one-dimensional Brownian motion;
- If $i \neq j$, then the processes W_t^i and W_t^j are independent.

Associated with a d-dimensional Brownian motion, we have a filtration $(\mathcal{F}_t)_{t \geq 0}$ such that:

- For each t, the random vector W_t is $\mathcal{F}_t$-measurable;
- For each $t \leq t_1 \leq \ldots \leq t_n$, the vector increments
$$W_{t_1} - W_t, \ldots, W_{t_n} - W_{t_{n-1}}$$
are independent of $\mathcal{F}_t$.

11.4.1 Cross-variations of Brownian motions

Because each component W_t^i of W_t is a one-dimensional Brownian motion, we have

$$[W^i]_t = t, \quad i = 1, \ldots, d.$$

However, if we define the cross-variation between W^i and W^j as

$$[W^i, W^j]_t := \lim_{\|\mathcal{P}\| \to 0} \sum_{k=0}^{n-1} (W^i_{t_{k+1}} - W^i_{t_k})(W^j_{t_{k+1}} - W^j_{t_k}), \quad i, j = 1, \ldots, d,$$

where $\mathcal{P} = \{t_0, t_1, \ldots, t_n\}$ is a partition of $[0, t]$, then we have:

Theorem 11.14. *If $i \neq j$, then*
$$[W^i, W^j]_t = 0.$$

Proof. Let $\mathcal{P} = \{t_0, t_1, \ldots, t_n\}$ be a partition of $[0, t]$. For $i \neq j$, define the *sample cross variation* of W_t^i and W_t^j on $[0, t]$ to be

$$C_{\mathcal{P}} := \sum_{k=0}^{n-1} (W^i_{t_{k+1}} - W^i_{t_k})(W^j_{t_{k+1}} - W^j_{t_k}).$$

The increments appearing on the RHS of the above equation are all independent of one another and all have mean zero. Therefore

$$\mathbb{E}[C_{\mathcal{P}}] = 0.$$

We compute $\operatorname{var}(C_{\mathcal{P}}) = \mathbb{E}[C_{\mathcal{P}}^2]$. First note that

$$\begin{aligned} C_{\mathcal{P}}^2 &= \sum_{k=0}^{n-1} (W^i_{t_{k+1}} - W^i_{t_k})^2 (W^j_{t_{k+1}} - W^j_{t_k})^2 \\ &+ 2 \sum_{\ell < k}^{n-1} (W^i_{t_{\ell+1}} - W^i_{t_\ell})(W^j_{t_{\ell+1}} - W^j_{t_\ell})(W^i_{t_{k+1}} - W^i_{t_k})(W^j_{t_{k+1}} - W^j_{t_k}). \end{aligned}$$

All the increments appearing in the sum of cross terms are independent of one another and have mean zero. Therefore

$$\mathrm{var}(C_{\mathcal{P}}) = \mathbb{E}[C_{\mathcal{P}}^2] = \sum_{k=0}^{n-1} (W^i_{t_{k+1}} - W^i_{t_k})^2 (W^j_{t_{k+1}} - W^j_{t_k})^2.$$

But $(W^i_{t_{k+1}} - W^i_{t_k})^2$ and $(W^j_{t_{k+1}} - W^j_{t_k})^2$ are independent of one another, and each has expectation $(t_{k+1} - t_k)$. It follows that

$$\mathrm{var}(C_{\mathcal{P}}) = \sum_{k=0}^{n-1} (t_{k+1} - t_k)^2 \le \|\mathcal{P}\| \sum_{k=0}^{n-1} (t_{k+1} - t_k) = \|\mathcal{P}\| t.$$

As $\|\mathcal{P}\| \to 0$ we have $\mathrm{var}(C_{\mathcal{P}}) \to 0$ so $C_{\mathcal{P}}$ converges in mean square[7] to the constant $\mathbb{E}[C_{\mathcal{P}}] = 0$.

□

11.4.2 Lévy's characterisation of Brownian motion*

Lévy's characterisation of BM (as given in Theorem 9.18 extends to the multi-dimensional case (see Shreve [13], Section 4.6.3 for more details).

Theorem 11.15 (Lévy's theorem, d-dimensional)**.** *Let M be a d-dimensional martingale relative to a filtration, with $M_0 = 0$, continuous paths, and $[M^i, M^j]_t = \delta^{ij} t$ for all $t \ge 0$. Then M is a d-dimensional BM.*

11.5 Two-dimensional Itô formula

There is a multi-dimensional version of the Itô formula. We content ourselves for now with the following two-dimensional version. The formula generalises (as we shall see) to *any number* of processes driven by a Brownian motion of *any number* (not necessarily the same number) of dimensions. Let $W := (W^1, W^2)$ be a two-dimensional Brownian motion (so that W^1, W^2 are independent Brownian motions), and let $X := (X^1, X^2)$ be a two-dimensional Itô process following

$$\mathrm{d}X_t = a_t\,\mathrm{d}t + b_t\,\mathrm{d}W_t, \tag{11.5}$$

where

$$a_t = \begin{pmatrix} a^1_t \\ a^2_t \end{pmatrix} \quad b_t = \begin{pmatrix} b^{11}_t & b^{12}_t \\ b^{21}_t & b^{22}_t, \end{pmatrix}$$

so that (11.5) is equivalent to

$$\begin{aligned} \mathrm{d}X^1_t &= a^1_t\,\mathrm{d}t + b^{11}_t\,\mathrm{d}W^1_t + b^{12}_t\,\mathrm{d}W^2_t, \\ \mathrm{d}X^2_t &= a^2_t\,\mathrm{d}t + b^{21}_t\,\mathrm{d}W^1_t + b^{22}_t\,\mathrm{d}W^2_t, \end{aligned}$$

[7] The convergence also holds almost surely, though we do not prove this here.

or in integral form

$$\begin{aligned} X_t^1 &= x^1 + \int_0^t a_s^1 \,\mathrm{d}s + b_s^{11}\,\mathrm{d}W_s^1 + b_s^{12}\,\mathrm{d}W_s^2, \\ X_t^2 &= x^2 + \int_0^t a_s^2 \,\mathrm{d}s + b_s^{21}\,\mathrm{d}W_s^1 + b_s^{22}\,\mathrm{d}W_s^2, \end{aligned}$$

or in compact form

$$X_t = x + \int_0^t a_s \,\mathrm{d}s + b_s \,\mathrm{d}W_s, \quad t \geq 0, \quad x = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}.$$

Such processes, consisting of a nonrandom initial condition, plus a Riemann integral, plus one or more Itô integrals, are examples of *semimartingales*. The integrands a_s, b_s can be any adapted processes such that the relevant integrals exist. The adaptedness of the integrands guarantees that X is also adapted.

Theorem 11.16 (Two-dimensional Itô formula)**.** *Let $f(t, x_1, x_2)$ be a function $f : [0,\infty) \times \mathbb{R}^2 \to \mathbb{R}$. Then the process $Y := (Y_t)_{t\geq 0}$ defined by $Y_t := f(t, X_t^1, X_t^2) \equiv f(t, X_t)$ follows*

$$\begin{aligned} \mathrm{d}Y_t &= f_t(t, X_t^1, X_t^2)\,\mathrm{d}t + f_{x_1}(t, X_t^1, X_t^2)\,\mathrm{d}X_t^1 + f_{x_2}(t, X_t^1, X_t^2)\,\mathrm{d}X_t^2 \\ &+ \frac{1}{2} f_{x_1x_1}(t, X_t^1, X_t^2)\,\mathrm{d}[X^1]_t + \frac{1}{2} f_{x_2x_2}(t, X_t^1, X_t^2)\,\mathrm{d}[X^2]_t + f_{x_1x_2}(t, X_t^1, X_t^2)\,\mathrm{d}[X^1, X^2]_t, \end{aligned}$$

where $\mathrm{d}[X^i, X^j]_t = \mathrm{d}X_t^i\,\mathrm{d}X_t^j, i = 1, 2$ *are computed according to the rules*

$$\mathrm{d}t\,\mathrm{d}t = \mathrm{d}t\,\mathrm{d}W_t^i = \mathrm{d}W_t^i\,\mathrm{d}t = 0, \quad \mathrm{d}W_t^i\,\mathrm{d}W_t^j = \delta^{ij}\,\mathrm{d}t,$$

with

$$\delta^{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

In integral form the theorem is

$$\begin{aligned} Y_t - Y_0 &= f(t, X_t^1, X_t^2) - f(0, X_0^1, X_0^2) \\ &= \int_0^t f_t(s, X_s^1, X_s^2)\,\mathrm{d}s + \int_0^t f_{x_1}(s, X_s^1, X_s^2)\,\mathrm{d}X_s^1 + \int_0^t f_{x_2}(s, X_s^1, X_s^2)\,\mathrm{d}X_s^2 \\ &+ \frac{1}{2}\int_0^t f_{x_1x_1}(s, X_s^1, X_s^2)\,\mathrm{d}[X^1]_s + \frac{1}{2}\int_0^t f_{x_2x_2}(s, X_s^1, X_s^2)\,\mathrm{d}[X^2]_s \\ &+ \int_0^t f_{x_1x_2}(s, X_s^1, X_s^2)\,\mathrm{d}[X^1, X^2]_s. \end{aligned}$$

11.5.1 Markovian diffusion case

If, in (11.5), we have $a_t = a(t, X_t), b_t = b(t, X_t)$ for well-behaved[8] functions $a(t, x), b(t, x)$, so that

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}t + b(t, X_t)\,\mathrm{d}W_t,$$

[8] Lipschitz continuity and linear growth conditions are usually sufficient.

then the process X is Markovian:

$$\mathbb{E}[h(X_T)|\mathcal{F}_t] = \mathbb{E}[h(X_T)|X_t], \quad 0 \le t \le T.$$

The integral equation for Y may be written

$$Y_t = Y_0 + \int_0^t (f_t(s, X_s) + \mathcal{A}f(s, X_s))\, \mathrm{d}s + \int_0^t (\nabla f(s, X_s))^{\mathrm{T}} b(s, X_s)\, \mathrm{d}W_s,$$

where $\mathcal{A}$ is the *generator* of the two-dimensional diffusion X, defined by

$$\begin{aligned} \mathcal{A}f(t,x) \quad &:= \quad \sum_{i=1}^{2} a^i(t,x) f_{x_i}(t,x) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} (bb^*)^{ij}(t,x) f_{x_i x_j}(t,x) \\ &= \quad a^*(t,x)\nabla f(t,x) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} (bb^*)^{ij}(t,x) f_{x_i x_j}(t,x), \end{aligned} \tag{11.6}$$

where * denotes matrix transposition, and where

$$\nabla f(t,x) = \begin{pmatrix} f_{x_1}(t,x) \\ f_{x_1}(t,x) \end{pmatrix}.$$

Exercise 11.17 (The product rule). Let X^1, X^2 be Itô processes:

$$\begin{aligned} X_t^1 &= x^1 + \int_0^t a_s^1\, \mathrm{d}s + \int_0^t b^{11}(s)\, \mathrm{d}W_s^1 + \int_0^t b_s^{12}\, \mathrm{d}W_s^2, \\ dX_t^2 &= \int_0^t a_s^2\, \mathrm{d}s + \int_0^t b_s^{21}\, \mathrm{d}W_s^1 + \int_0^t b_s^{22}\, \mathrm{d}W_s^2. \end{aligned}$$

Use a two-dimensional Itô formula to derive the *product rule*

$$\mathrm{d}(X_t^1 X_t^2) = X_t^1\, \mathrm{d}X_t^2 + X_t^2\, \mathrm{d}X_t^1 + \mathrm{d}X_t^1\, \mathrm{d}X_t^2,$$

or, in integral form

$$[X^1, X^2]_t = \int_0^t X_s^1\, \mathrm{d}X_s^2 + \int_0^t X_s^2\, \mathrm{d}X_s^1 + \int_0^t \mathrm{d}[X^1, X^2]_s.$$

11.6 Multidimensional Itô formula

11.6.1 Multidimensional Itô process

Let $W_t = (W_t^1, \ldots, W_t^d)$ be a vector of d independent Brownian motions, that is, W_t is d-dimensional Brownian motion. We can use the Brownian motion vector to form the following n Itô processes $X_t^1, \ldots, X_t^n$:

$$\begin{aligned} \mathrm{d}X_t^1 \quad &= \quad a_t^1\, \mathrm{d}t + b_t^{11}\, \mathrm{d}W_t^1 + \cdots + b_t^{1d}\, \mathrm{d}W_t^d \\ &\vdots \\ dX_t^n \quad &= \quad a_t^n\, \mathrm{d}t + b_t^{n1}\, \mathrm{d}W_t^1 + \cdots + b_t^{nd}\, \mathrm{d}W_t^d, \end{aligned}$$

or, in matrix notation, with $X = (X^1, \dots, X^n)^*$,

$$\mathrm{d}X_t = a_t\,\mathrm{d}t + b_t\,\mathrm{d}W_t, \tag{11.7}$$

where

$$X_t = \begin{pmatrix} X_t^1 \\ \vdots \\ X_t^n \end{pmatrix} \quad a_t = \begin{pmatrix} a_t^1 \\ \vdots \\ a_t^n \end{pmatrix} \quad b_t = \begin{pmatrix} b_t^{11} & \cdots & b_t^{1d} \\ \vdots & \vdots & \vdots \\ b_t^{n1} & \cdots & b_t^{nd} \end{pmatrix} \tag{11.8}$$

Note that the coefficients a and b are required to satisfy certain conditions so that the integrals implicit in the above equations are well defined. In particular, their elements should all be adapted process, so that we know their values at time t if we know X_t.

Theorem 11.18 (Multidimensional Itô formula)**.** *Suppose X_t satisfies (11.7). Let*

$$f(t,x) = (f_1(t,x), \dots, f_p(t,x))^*$$

be a twice differentiable map from $[0,\infty) \times \mathbb{R}^n$ into $\mathbb{R}^p$. Then the process $Y_t := f(t, X_t)$ is again an Itô process, whose k^{th} component, Y_t^k, is given by the multidimensional Itô formula as

$$\mathrm{d}Y_t^k = \frac{\partial f_k}{\partial t}(t,X_t)\,\mathrm{d}t + \sum_{i=1}^n \frac{\partial f_k}{\partial x_i}(t,X_t)\,\mathrm{d}X_t^i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 f_k}{\partial x_i \partial x_j}(t,X_t)\,\mathrm{d}[X^i, X^j]_t, \tag{11.9}$$

where $\mathrm{d}[X^i, X^j]_t \equiv \mathrm{d}X_t^i\,\mathrm{d}X_t^j$ is computed according to the rules

$$\mathrm{d}W_t^i\,\mathrm{d}W_t^j = \delta_{ij}\,\mathrm{d}t, \quad \mathrm{d}t\,\mathrm{d}t = \mathrm{d}W_t^i\,\mathrm{d}t = \mathrm{d}t\,\mathrm{d}W_t^i = 0.$$

Example 11.19. Let $W = (W^1, \dots, W^n)$ be Brownian motion in $\mathbb{R}^n$, for $n \geq 2$. Consider

$$R_t := |W_t| = ((W^1)_t^2 + \cdots + (W^n)_t^2)^{1/2},$$

which is a process describing the distance of the n-dimensional Brownian motion from the origin. Now, the function $f(t,x) = |x|$ is not differentiable at the origin, but since W_t never hits the origin (almost surely, or with probability one) when $n \geq 2$ (see, for example, Øksendal [11], Exercise 9.7), the multidimensional Itô formula still works.

Take $X_t = W_t$, so that $\mathrm{d}X_t = \mathrm{d}W_t$, and consider the process $Y_t = R_t = f(t, X_t) = f(t, W_t) = |W_t| = ((W_t^1)^2 + \cdots + (W_t^n)^2)^{1/2}$. Then $f(t,x) = (x_1^2 + \ldots + x_n^2)^{1/2}$, so that

$$\frac{\partial f}{\partial t} = 0, \quad \frac{\partial f}{\partial x_i} = \frac{x_i}{|x|}, \quad \frac{\partial^2 f}{\partial x_i^2} = \frac{\delta_{ij}}{|x|} - \frac{x_i x_j}{|x|^3}.$$

Then

$$\begin{aligned} \mathrm{d}R_t &= \sum_{i=1}^n \frac{W_t^i}{|W_t|}\,\mathrm{d}W_t^i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \left(\frac{\delta_{ij}}{|W_t|} - \frac{W_t^i W_t^j}{|W_t|^3}\right)\delta_{ij}\,\mathrm{d}t \\ &= \sum_{i=1}^n \frac{W_t^i\,\mathrm{d}W_t^i}{R_t} + \frac{n-1}{2R_t}\,\mathrm{d}t. \end{aligned}$$

11.7 Connections with PDEs: Feynman-Kac theorem

There is a remarkable connection between stochastic calculus for Markov diffusions and partial differential equations (PDEs). Consider the one-dimensional diffusion

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}t + b(t, X_t)\,\mathrm{d}W_t. \tag{11.10}$$

The process $X = (X_t)_{t\geq 0}$ is a Markov process, satisfying

$$\mathbb{E}[h(X_T)|\mathcal{F}_t] = \mathbb{E}[h(X_T)|X_t], \quad 0 \leq t \leq T, \tag{11.11}$$

for a function $h(x)$ such that the above expectations are defined. A consequence of the Markov property is that the right-hand-side of (11.11) is a function of (t, X_t) only. Write

$$v(t, x) := \mathbb{E}[h(X_T)|X_t = x]. \tag{11.12}$$

Lemma 11.20. *The process $Y = (Y_t)_{0\leq t\leq T}$ defined by $Y_t := v(t, X_t)$ is a martingale.*

Proof. By the Markov property, we have $Y_t = \mathbb{E}[h(X_T)|X_t] = \mathbb{E}[h(X_T)|\mathcal{F}_t]$. Then, for $0 \leq s \leq t \leq T$,

$$\begin{aligned}
\mathbb{E}[Y_t|\mathcal{F}_s] &= \mathbb{E}[\mathbb{E}[h(X_T)|X_t]\mathcal{F}_s] \\
&= \mathbb{E}[\mathbb{E}[h(X_T)|\mathcal{F}_t]|\mathcal{F}_s] \quad \text{(by Markov property)} \\
&= \mathbb{E}[h(X_T)|\mathcal{F}_s] \quad \text{(by Tower property)} \\
&= \mathbb{E}[h(X_T)|X_s] \\
&= Y_s.
\end{aligned}$$

□

Theorem 11.21 (Feynman-Kac)**.** *The function $v(t, x)$ in (11.12) satisfies the PDE*

$$v_t(t, x) + a(t, x)v_x(t, x) + \frac{1}{2}b^2(t, x)v_{xx}(t, x) = 0, \quad v(T, x) = h(x). \tag{11.13}$$

Proof. By the Itô formula

$$\begin{aligned}
\mathrm{d}Y_t &= \mathrm{d}v(t, X_t) \\
&= [v_t(t, X_t) + a(t, X_t)v_x(t, X_t) + \frac{1}{2}b^2(t, X_t)v_{xx}(t, X_t)]\,\mathrm{d}t + b(t, X_t)v_x(t, X_t)\,\mathrm{d}W_t.
\end{aligned}$$

Since Y is a martingale the coefficient of the "$\mathrm{d}t$" term must be zero for all (t, X_t), and (11.13) follows.

□

Note that the PDE (11.13) may be written

$$v_t(t, x) + \mathcal{A}v(t, x) = 0, \quad v(T, x) = h(x), \tag{11.14}$$

where $\mathcal{A}$ is the generator of the diffusion (11.10):

$$\mathcal{A}v(t,x) = a(t,x)v_x(t,x) + \frac{1}{2}b^2(t,x)v_{xx}(t,x),$$

(and this is the form that generalises to a multi-dimensional situation). Note also that the theorem is still valid if we replace $h(X_T)$ in (11.12) by $h(T, X_T)$, a function dependent on T as well as X_T.

Finally, there is an obvious generalisation to a multi-dimensional situation. We content ourselves with the following two-dimensional version.

Suppose we have a two-dimensional diffusion $X = (X^1, X^2)$ following

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}t + b(t, X_t)\,\mathrm{d}W_t, \tag{11.15}$$

where

$$a(t, X_t) = \begin{pmatrix} a^1(t, X_t) \\ a^2(t, X_t) \end{pmatrix}, \quad b(t, X_t) = \begin{pmatrix} b^{11}(t, X_t) & b^{12}(t, X_t) \\ b^{21}(t, X_t) & b^{22}(t, X_t), \end{pmatrix}$$

so that (11.15) is equivalent to

$$\begin{aligned} \mathrm{d}X_t^1 &= a^1(t, X_t)\,\mathrm{d}t + b^{11}(t, X_t)\,\mathrm{d}W_t^1 + b^{12}(t, X_t)\,\mathrm{d}W_t^2, \\ \mathrm{d}X_t^2 &= a^2(t, X_t)\,\mathrm{d}t + b^{21}(t, X_t)\,\mathrm{d}W_t^1 + b^{22}(t, X_t)\,\mathrm{d}W_t^2. \end{aligned}$$

Let $h(t,x) \equiv h(t, x_1, x_2)$ be a function $h : [0,\infty) \times \mathbb{R}^2 \to \mathbb{R}$. Define the function

$$v(t,x) := \mathbb{E}[h(X_T)|X_t = x].. \tag{11.16}$$

The generator of the diffusion (11.15) is $\mathcal{A}$, given by (11.6):

$$\mathcal{A}f(t,x) := \sum_{i=1}^{2} a^i(t,x) f_{x_i}(t,x) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} (bb^*)^{ij}(t,x) f_{x_i x_j}(t,x).$$

Theorem 11.22 (Feynman-Kac, two-dimensional)**.** *The function $v(t,x)$ in (11.16) satisfies the PDE*

$$v_t(t,x) + \mathcal{A}v(t,x) = 0, \quad v(T,x) = h(x),$$

where $\mathcal{A}$ is the generator of the diffusion (11.15).

11.8 The Girsanov Theorem*

Given a Brownian motion $W := (W_t)_{0 \le t \le T}$ on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with the filtration $\mathbb{F} := (\mathcal{F}_t)_{0 \le t \le T}$ being that generated by W, and given an adapted process $\theta := (\theta_t)_{0 \le t \le T}$, define the (local) martingale Z by

$$Z_t := \mathcal{E}(-\theta \cdot W)_t := \exp\left(-\int_0^t \theta_s\,\mathrm{d}W_s - \frac{1}{2}\int_0^t \theta_s^2\,\mathrm{d}s\right), \quad 0 \le t \le T,$$

where $\mathcal{E}$ is the so-called Doléans exponential. We have that Z follows

$$\mathrm{d}Z_t = -\theta_t Z_t \,\mathrm{d}W_t.$$

Then, provided θ satisfies the *Novikov condition*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \theta_t^2 \,\mathrm{d}t\right)\right] < \infty, \tag{11.17}$$

we can define a new probability measure $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F} \equiv \mathcal{F}_T$ by

$$\mathbb{Q}(A) = \int_A Z_T \,\mathrm{d}\mathbb{P}, \quad \forall A \in \mathcal{F},$$

and the process

$$W_t^{\mathbb{Q}} := W_t + \int_0^t \theta_s \,\mathrm{d}s, \quad 0 \le t \le T,$$

is a $\mathbb{Q}$-Brownian motion. We write $Z_T = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ and we have, for any $\mathcal{F}$-measurable random variable X,

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}[XZ_T]. \tag{11.18}$$

Remark 11.23. The Novikov condition (11.17) is sufficient to guarantee that Z is a $(\mathbb{P}, \mathbb{F})$-martingale, so that $\mathbb{E}[Z_T] = 1$ and $\mathbb{Q}$ is indeed a probability measure.

As well as (11.18) we have the following results connecting conditional expectations under $\mathbb{Q}$ and $\mathbb{P}$.

Let $0 \le t \le T$. If X is $\mathcal{F}_t$-measurable, then

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}[XZ_t].$$

Bayes formula If X is $\mathcal{F}_t$-measurable and $0 \le s \le t \le T$, then

$$Z_s \mathbb{E}^{\mathbb{Q}}[X|\mathcal{F}_s] = \mathbb{E}[XZ_t|\mathcal{F}_s].$$

There is a multi-dimensional version of Girsanov's Theorem. Once again we content ourselves with a two-dimensional version. Given a two-dimensional Brownian motion $W = (W^1, W^2)$ on a stochastic basis $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P})$, and a two-dimensional adapted process $\theta = (\theta^1, \theta^2)$, define a (local) martingale Z by

$$\begin{aligned} Z_t &= \mathcal{E}(-\theta \cdot W)_t \equiv \mathcal{E}(-\theta^1 \cdot W^1 - \theta^2 \cdot W^2)_t \\ &= \exp\left(-\int_0^t \theta_s^1 \,\mathrm{d}W_s^1 - \int_0^t \theta_s^2 \,\mathrm{d}W_s^2 - \frac{1}{2}\int_0^t \left[(\theta_s^1)^2 + (\theta_s^2)^2\right] \mathrm{d}s\right). \end{aligned}$$

Then, provided we have the two-dimensional Novikov condition

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \left[(\theta_t^1)^2 + (\theta_t^2)^2\right] \mathrm{d}s\right)\right] < \infty, \tag{11.19}$$

we can define a new probability measure $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F} \equiv \mathcal{F}_T$ by

$$\mathbb{Q}(A) = \int_A Z_T \,\mathrm{d}\mathbb{P}, \quad \forall A \in \mathcal{F},$$

and the process $W^{\mathbb{Q}} = (W^{\mathbb{Q},1}, W^{\mathbb{Q},2})$ defined by

$$W_t^{\mathbb{Q},1} := W_t + \int_0^t \theta_s^1 \,\mathrm{d}s, \quad W_t^{Q,2} := W_t + \int_0^t \theta_s^2 \,\mathrm{d}s.$$

is a two-dimensional $\mathbb{Q}$-Brownian motion.

References

[1] T. Björk, *Arbitrage theory in continuous time*, Oxford University Press, third ed., 2009.

[2] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Polit. Econ., 81 (1973), pp. 637–659.

[3] M. H. A. Davis and A. Etheridge, *Louis Bachelier's* Theory of Speculation*: the origins of modern finance*, PUP, 2006.

[4] A. Etheridge, *A course in financial calculus*, Cambridge University Press, Cambridge, 2002.

[5] G. R. Grimmett and D. R. Stirzaker, *Probability and random processes*, Oxford University Press, New York, third ed., 2001.

[6] J. M. Harrison and S. R. Pliska, *Martingales and stochastic integrals in the theory of continuous trading*, Stochastic Process. Appl., 11 (1981), pp. 215–260.

[7] J. C. Hull, *Options, futures and other derivatives*, Pearson, eighth ed., 2011.

[8] J. Jacod and P. Protter, *Probability essentials*, Universitext, Springer-Verlag, Berlin, second ed., 2003.

[9] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, vol. 113 of Graduate Texts in Mathematics, Springer-Verlag, New York, second ed., 1991.

[10] R. C. Merton, *Theory of rational option pricing*, Bell J. Econom. and Management Sci., 4 (1973), pp. 141–183.

[11] B. Øksendal, *Stochastic differential equations*, Universitext, Springer-Verlag, Berlin, sixth ed., 2003. An introduction with applications.

[12] S. E. Shreve, *Stochastic calculus for finance. I*, Springer Finance, Springer-Verlag, New York, 2004. The binomial asset pricing model.

[13] ——, *Stochastic calculus for finance. II*, Springer Finance, Springer-Verlag, New York, 2004. Continuous-time models.

[14] D. Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.

[15] P. Wilmott, S. Howison, and J. Dewynne, *The mathematics of financial derivatives*, Cambridge University Press, Cambridge, 1995. A student introduction.